POLITECNICO MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE E DELL'INFORMAZIONE

EXECUTIVE SUMMARY OF THE THESIS

Integrating machine learning and derivative-free optimizers for oil production optimization by waterflooding

LAUREA MAGISTRALE IN ENERGY ENGINEERING - INGEGNERIA ENERGETICA

Author: GUIDO DI FEDERICO

Advisor: PROF. ENRICO ZIO

Co-advisors: DR. AHMED SHOKRY, DR. GIORGIO FIGHERA, DR. EMANUELE VIGNATI

Academic year: 2020-2021

---

1. Introduction

The recent rise in global energy demand and the decline in new substantial discoveries require a careful management of existing oil and gas fields. After a first stage of primary production, driven by the pressure of reservoir fluids, the most common method to enhance recovery is injecting water from wells (i.e. waterflooding). Waterflooding optimization consists in identifying the optimal injection scheme for a given field, to maximize its economic production and reduce wastes. While traditional surveillance methods can lead to suboptimal production strategies, more complex simulation-based optimization is often computationally heavy, in terms of the resources required to construct, tune and run full-physics simulations. Hence, recent literature [1–3] has focused on the use of surrogate-based optimization (SBO), in which the simulator is replaced by a data-driven model, requiring only production data, without any geological information. While full-physics models are multi-purpose, surrogate models are typically developed for the specific problem.

2. Problem statement

The generic field is made up of N_{in} injection wells, with injection rates $q_{in,i}$ and N_p production wells with water and oil production rates $q_{w,j}$ and $q_{o,j}$, located at specific positions, as in Figure 1. Additionally, the problem is subject to the set of constraints $\boldsymbol{c}$, which model the technical and operational limits of the field, such as injection rate bounds for each well and field injection rate limit (inequality constraint).

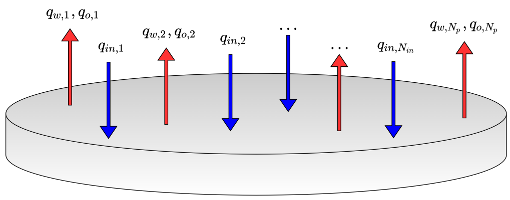

Figure 1: Schematic field and well designation.

The generic problem of waterflooding optimization aims at finding the control variables $\boldsymbol{u}$ which result in the optimal value of the objective function J, while satisfying constraints $\boldsymbol{c}$:

$$\begin{cases} \min_{\boldsymbol{u} \in U} J(\boldsymbol{x}, \boldsymbol{u}) \\ \boldsymbol{c}(\boldsymbol{x}, \boldsymbol{u}) \leqslant \boldsymbol{0} \\ \boldsymbol{g}(\boldsymbol{x}, \boldsymbol{u}) = \boldsymbol{0} \end{cases}$$

where:

- $\boldsymbol{x}$ is the vector of dynamic state variables of the model (pressure, saturation, etc.);
- $\boldsymbol{u}$ is the vector of well control variables, of dimension n (i.e. injection rates);
- $U = \{\boldsymbol{u} \in \mathbb{R}^n; \boldsymbol{u}_{lb} \leqslant \boldsymbol{u} \leqslant \boldsymbol{u}_{ub}\}$ defines the allowable values for $\boldsymbol{u}$;
- $\boldsymbol{c}$ is the set of linear and nonlinear constraints on all control variables;
- $\boldsymbol{g}$ is the reservoir model (set of reservoir simulation equations) to be solved to evaluate J and $\boldsymbol{c}$.

The most common objective function for waterflooding optimization problems is the net present value (NPV) [4]. It gives an economic evaluation of the field's performance in terms of costs and revenues. Since the NPV is to be maximized, $J(\boldsymbol{u})$ = -NPV($\boldsymbol{u}$). Mathematically, the NPV can be defined as:

$$\frac{\sum_{k=1}^{N_t} \left[\sum_{j=1}^{N_p} (r_o q_{o,j}^{k\Delta t} - r_w q_{w,j}^{k\Delta t}) - \sum_{i=1}^{N_{in}} r_{in} q_{in,i}^{k\Delta t} \right] \Delta t}{(1+d)^{\frac{k\Delta t}{365}}}$$

where r_o is the price of produced oil per unit volume, r_w is the cost of produced water per unit volume, r_{in} is the cost of injected water per unit volume and d is the yearly discount factor (such that if time steps are expressed in days the ratio $\frac{k\Delta t}{365}$ is dimensionless).

3. Methodology

This thesis presents an innovative SBO framework for wateflooding management of mature brownfields, which integrates machine learning (ML) models, such as long short-term memory (LSTMs) or physics informed neural networks (PINNs), with optimization techniques. The proposed framework consists of three stages.

The first stage is data collection, i.e. water injection rates and corresponding water and oil production rates. In case a reservoir model is available, this stage can be improved by simulating input-outpupt data for a variety of production strategies. In this thesis, since the methodology is applied to synthetic reservoir cases, realistic data generation through simulation of the historical period represents the data collection phase. The second stage includes the development of a ML model (PINN or LSTM) which is able to accurately predict the future oil and water production rates at each well as a function of the water injection rates. In the third stage the developed ML model is coupled to an optimization algorithm to identify the optimal water injection profiles to be applied to the field over the future time period, with the NPV as objective function. Three different optimization algorithms, including ensemble-based (EnOpt, see Chen et al. (2009)[5]), genetic, and gradient-based (trust-region), are used to evaluate the flexibility and robustness of the framework. The obtained optimal solutions are then compared to standard reservoir practices (do-nothing and pressure maintenance scenarios, called VR) and software (FloodOpt® by *StreamSim*®, which uses a heursitic optimization algorithm based on injection efficiency) in terms of objective function, computational time, and optimization strategies. These scenarios are only intended as a baseline for comparison, since they require a tuned geological model to be evaluated. FloodOpt® allows to optimize production with a given injection target, but it does not allow to specify an explicit objective function. Thus, for a fair comparison, the best-performing algorithm is adapted by modifying the objective function with oil produced only (non-discounted) and by using a target injection at a field level (equality constraint). This case, which is referred as "target", is performed for validation purposes only: the increased flexibility of the proposed methodology is, in real life, a benefit to take advantage of.

With respect to the approaches developed in literature, the novelty of the proposed framework lies in:

- using PINNs and LSTMs in the context of injection schedule optimization of mature fields
- presenting a comparison with a state-of-the-art commercial software to understand if time and effort required for a 3D reservoir model are justified or if a data-driven approach is more convenient
- the deep interpretation and verifications of the obtained optimal solutions
- the use of EnOpt [5] in SBO of oil field pro-

duction optimization

4. Application

The effectiveness of the proposed framework is validated through its application to two case studies, the Streak field and the Olympus field.

4.1. Streak field

The Streak field is a 2D reservoir model with homogeneous geological properties, including 4 production and 5 injection wells, as in Figure 2.

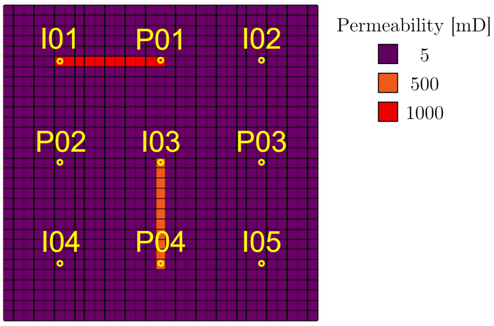

Figure 2: Streak: geological model.

Maniglio et al. (2021)[6] trained two types of networks for the Streak field, which are used in this work. The goal is to apply the optimization workflow to such surrogates. They both receive injection rates and times at all injectors as inputs: one is a traditional ANN and forecasts water cut, as shown in Figure 3, while the other is a PINN (ANN combined with a capacitance-resistance model) and forecasts liquid production rates at all producers, as shown in Figure 4.

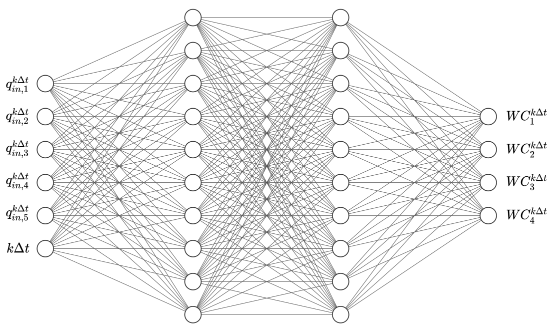

Figure 3: Streak: water cut ANN.

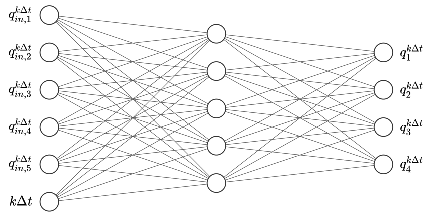

Figure 4: Streak: liquid production rate PINN.

Overall, all three algorithms show an increase in the objective function compared to the do-nothing case, although with different computational times. EnOpt reaches the highest increase and proves to be the fastest, while the GA the slowest, as reported in Figure 5. In general, the PINN-based forward model of the Streak reservoir achieves a significant reduction in the elapsed time for a single forward evaluation of the objective function, on the order of about 10 compared to Eclipse® commercial simulator.

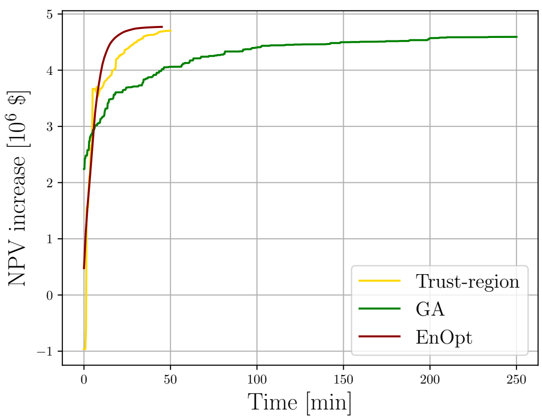

Figure 5: Streak: NPV increase compared to the do-nothing case for the three algorithms.

The optimization algorithms tend to balance between high oil production (higher profit) and low water injection and production (lower cost), given the high water cuts. In particular, while the GA and EnOpt achieve a higher cumulative oil production than the base case, with a significant decrease in the injected water, the trust-region algorithm improves the objective function by slightly reducing oil production, but injecting and producing the least water. The VR strategy gives a much lower NPV value compared to the do-nothing case because the total injection rate required to maintain reservoir pressure stable results in high costs with limited return. Cumulatives are shown in Table 1.

Cumulatives [10^6 stb]			
Case	**Wat. inj.**	**Wat. prod.**	**Oil prod.**
Trust-r.	4.02	3.98	0.102
GA	4.49	4.43	0.115
EnOpt	4.41	4.35	0.114
VR	7.57	7.49	0.108
Do-n.	7.06	6.97	0.112

Table 1: Streak: cumulative injection and production for each simulated strategy.

For the best algorithm, EnOpt, the comparison with the state-of-the-art software gave positive results as well, as reported in Table 2.

Cumulatives [10^6 stb]			
Case	**Wat. inj.**	**Wat. prod.**	**Oil prod.**
EnOpt (target)	7.60	7.44	0.166
FloodOpt	7.60	7.45	0.143

Table 2: Streak: cumulative injection and production for each simulated strategy (FloodOpt comparison).

In this case, most of the increase in oil production, and better management of the waterflooding process as a whole, comes from the area around injection well I01, as shown in the oil saturation difference map in Figure 6.

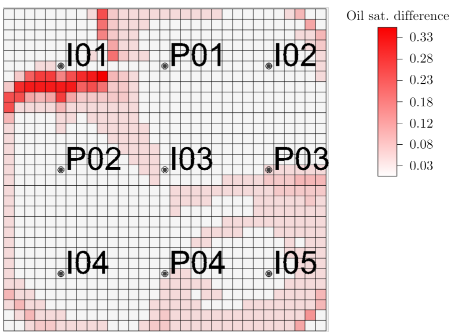

Figure 6: Streak: oil saturation difference with do-nothing after optimization (EnOpt (target)).

4.2. Olympus field

The Olympus field is a 3D reservoir model with complex geological properties, including 7 injection and 11 production wells, as in Figure 7 and, more in detail, in Figure 8.

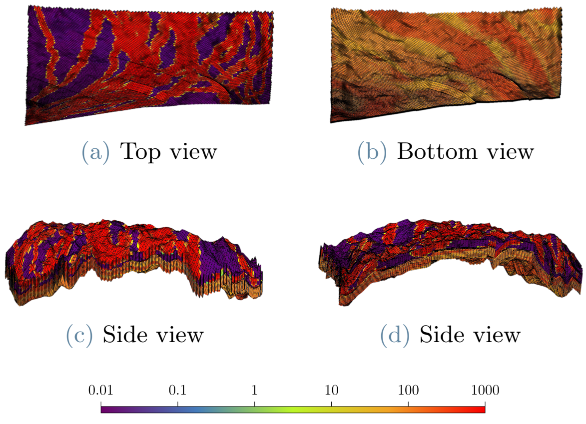

Figure 7: Olympus: geological model.

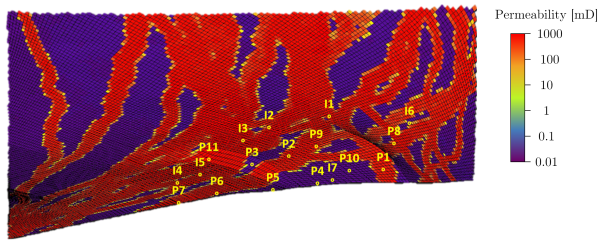

Figure 8: Olympus: top view highlighting well positions.

In the case of the Olympus field, the higher number of wells and the more complex geology require the explicit integration of time dependency into the surrogate model, differently from the Streak case. Two LSTM networks are trained for each production well, one for water and one for oil, for a total of 22 networks. They all receive injection rates at all injectors as inputs and forecast oil and water production rates at each producer, as shown in Figure 9.

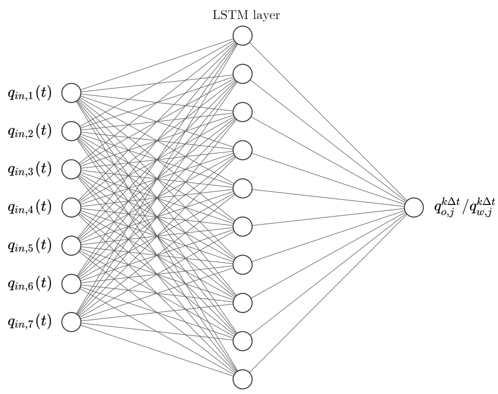

Figure 9: Olympus: generic LSTM network.

The Adam optimizer is used as training algorithm: more details can be found in Géron (2019)[7]. The hyperparameters for the networks are chosen based on a sensitivity analysis carried out using the values reported in Table 3 (where "past time steps" refers to the number of previous time steps used for forecasting by the LSTM).

Hyperparameter	Values
Hidden layers	1, 2
Neurons per layer	10, 20
Past time steps	3, 5
Learning rate	10^{-4}, 10^{-3}, 10^{-2}
L1 regularization	0, 10^{-2}, 10^{-1}
L2 regularization	10^{-3}, 10^{-2}, 10^{-1}

Table 3: Olympus: tested LSTM hyperparameters.

Overall, all three algorithms show an increase in the objective function, although with different computational times. EnOpt reaches the highest increase and proves again to be the fastest, while the GA the slowest, as reported in Figure 10. In general, the LSTM-based forward model of the Streak reservoir achieves a significant reduction in the elapsed time for a single forward evaluation of the objective function, on the order of about 100 compared to Eclipse® commercial simulator. In this case, the VR strategy still yields a better result than the do-nothing case, but lower compared to the algorithms.

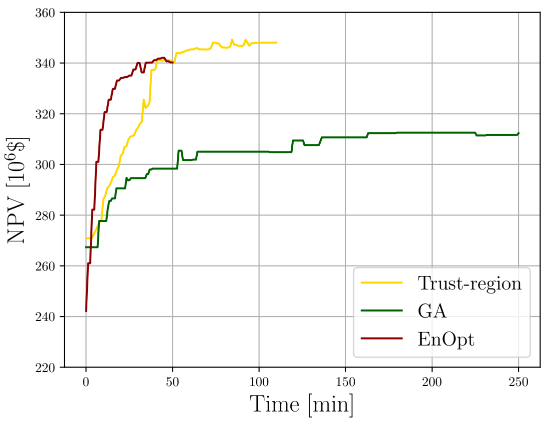

Figure 10: Olympus: NPV for the three algorithms.

Cumulatives are shown in Table 4.

	Cumulatives [10^6 Sm^3]		
Case	**Wat. inj.**	**Wat. prod.**	**Oil prod.**
Trust-r.	3.60	2.57	0.802
GA	3.56	2.53	0.798
EnOpt	3.55	2.55	0.802
VR	2.70	1.99	0.657
Do-n.	2.59	1.91	0.630

Table 4: Olympus: cumulative injection and production for each simulated strategy.

For the best-performing algorithm, EnOpt, the better management of the waterflooding process, which results in an increase in the objective function, can be better visualized with the aid of a mobile oil difference map, as in Figure 11. The green areas show regions where, at the end of the optimization process, there is less oil than for the do-nothing case (which is hopefully produced, or moved elsewhere), red areas where there is more (which is left underground). The cumulative sum of red and green areas is negative, confirming how EnOpt reaches a higher cumulative oil production.

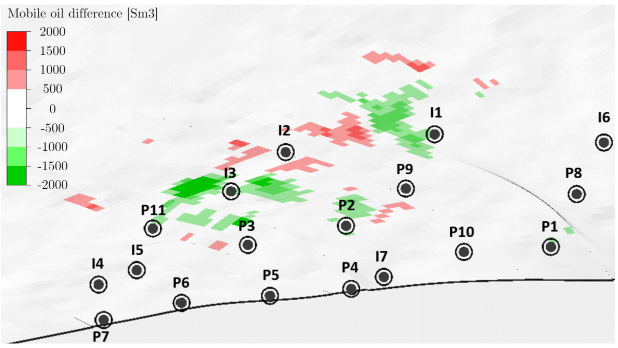

Figure 11: Olympus: mobile oil difference map (EnOpt-do-nothing)

The comparison of EnOpt (target) with the state-of-the-art software gave positive results as well, as reported in Table 5.

Cumulatives [10^6 Sm^3]			
Case	**Wat. inj.**	**Wat. prod.**	**Oil prod.**
EnOpt (target)	3.59	2.55	0.807
FloodOpt	3.59	2.52	0.799

Table 5: Olympus: cumulative injection and production for each simulated strategy (FloodOpt comparison).

For both case studies, optimized production scenarios are further analyzed with the aid of streamline maps, pressure and saturation distributions, to gain additional insights into the algorithms' optimization strategies and interpret results both from a mathematical and a reservoir engineering perspective.

5. Conclusions

The benefits of the proposed workflow confirm that it can be applied to the problem of waterflooding optimization of realistic brownfields, without the support of geological information. Simplifying assumptions include the timeline, where all wells share the same start date, the high quality of the datasets (neglecting noise, uncertainty or sparsity) and the constant dynamics of wells, which are not modified by workover operations. In particular, it is found that:

- surrogate models are able to approximate the behavior of the reservoir with good accuracy compared to the full-physics simulator, but with a significant decrease in the computational time;
- the optimization process leads to improved values of the objective function with respect to the "no-action" (do-nothing) and "pressure maintenance" (VR) scenarios and equivalent oil production to the simulator-based optimization software, due to an efficient allocation of the injected water;
- the optimization process provides high-level operational guidelines for the field;
- the developed workflow provides flexibility in specifying any objective function and constraints, differently from the benchmark software.

Future developments can include integrating PINNs and LSTMs, exploiting additional input data during training and/or optimization (e.g. as bottom-hole pressures), enhancing the parallelization of the workflow, investigating further model uncertainty quantification, as well as the application to a real field.

6. Acknowledgements

My sincere thanks go to my company supervisors at Eni S.p.A., Dr. Giorgio Fighera and Dr. Emanuele Vignati, and to Professor Enrico Zio and Dr. Ahmed Shokry. Their insights, suggestions and encouragement made the past year a period of both academic and personal growth.

References

[1] Y.D. Kim and L. Durlofsky. A Recurrent Neural Network–Based Proxy Model for Well-Control Optimization with Nonlinear Output Constraints. *SPE Journal*, 26(04):1837–1857, 08 2021.

[2] L. Deng and Y. Pan. Machine-Learning-Assisted Closed-Loop Reservoir Management Using Echo State Network for Mature Fields under Waterflood. *SPE Reservoir Evaluation & Engineering*, 23(04):1298–1313, 11 2020.

[3] A. Yewgat et al. Deep-crm: A new deep learning approach for capacitance resistive models. In *ECMORXVII: 17 th European Conference On The Mathematics Of Oil Recovery*, Online, France, September 2020.

[4] Jian Hou et al. A review of closed-loop reservoir management. *Petroleum Science*, 12:114–128, 2015.

[5] Yan Chen et al. Efficient ensemble-based closed-loop production optimization. *Spe Journal*, 14:634–645, 2009.

[6] M. Maniglio et al. Physics Informed Neural Networks Based on a Capacitance Resistance Model for Reservoirs Under Water Flooding Conditions. In *Abu Dhabi International Petroleum Exhibition and Conference*, 11 2021. D021S027R001.

[7] A. Géron. *Hands-on machine learning with Scikit-Learn and TensorFlow : concepts, tools, and techniques to build intelligent systems.* O'Reilly Media, 2019.