

Generating hybrid models for reservoir simulation

Bochkarev Artem

Moscow Institute of Physics and Technology
Department of Control and Applied Mathematics
Chair of Intelligent Systems

Research advisor V. Strijov
Expert I. Sofronov

Moscow, SMR,
6 December 2016

Problem

Build a machine learning model of reservoir that can predict amount of water and oil extraction, depending on schedule of water injection and bottom hole pressure.

Model requirements

- forecast has to be precise — it should give the minimum of some predefined loss function;
- model shouldn't require big amount of time and resources to predict extraction: the most consuming part has to be learning

Methods

We propose generating optimal models using machine learning techniques and a mix of real and synthetic data

- S. D. Mohaghegh *Reservoir simulation and modeling based on artificial intelligence and data mining*
- S. Koziel, L. Leifsson *Surrogate-Based Modeling and Optimization*
- Y. S. Ong, P. B. Nair, and A. J. Keane *Evolutionary Optimization of Computationally Expensive Problems via Surrogate Modeling*

Surrogate models

In this semester we started from building surrogate models of reservoir — models which use only synthetic data, which was precalculated by some existent simulation tool.

ECLIPSE

ECLIPSE is simulator, developed by Schlumberger, which provides numeric solution for reservoir modelling. The problem is ECLIPSE uses a lot of computational power and/or time to make single prediction.

We have injection (pump water inside reservoir) and producer wells. Let us denote time series from injection wells and pressure in producer wells as $\mathbf{x}^{(i)}$, where i - number of time series. Let us also denote the time series for oil extraction as \mathbf{y} .

Loss function

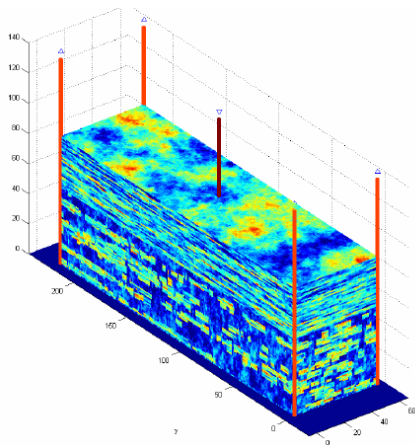
We want to find the model $f^* : X \rightarrow Y$ which minimizes the symmetric mean absolute percentage error (SMAPE):

$$f^* = \arg \min_f \frac{2}{n} \sum_{k=1}^N \frac{|y_k - \hat{y}_k|}{|y_k + \hat{y}_k|} \cdot 100\%$$

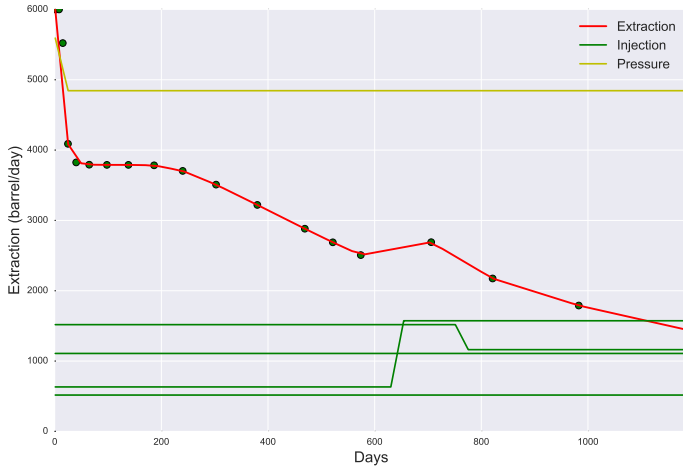
where $\mathbf{y} = f(\mathbf{x})$, $\hat{\mathbf{y}}$ — true value of time series.

Computational experiment

In the first computational experiment we worked with Eclipse data obtained from SPE10 dataset by Pavel Temrichev. We have 500 training and 300 testing samples, with 4 injection and 1 producer wells for each. Simulation was conducted for approximately 4 years.

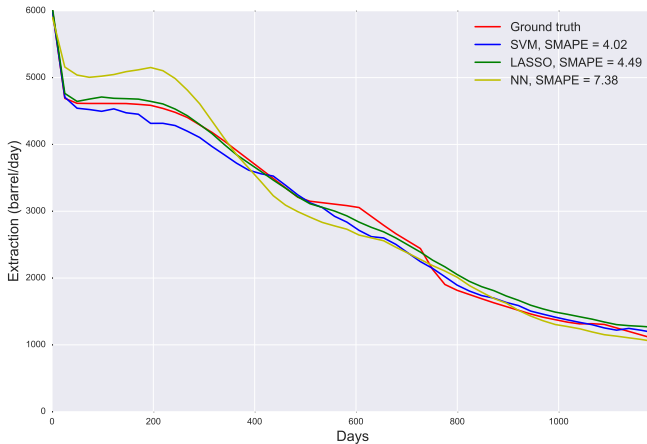


Computational experiment



Computational experiment

Model	LASSO	SVM	NN
SMAPE	6.35	6.24	11.03



Results

- Basic ECLIPSE knowledge and experience is obtained
- First surrogate model is created and tested

Project

- Improve accuracy of surrogate model
- Consider example of one more reservoir model
- Propose an approach to optimal data generation

Current activities

- Full-time Skoltech student
- MIPT education at CC RAS
- One day at Schlumberger