Data Driven Model Discovery – Petroleum application

Sander André Søndeland
University of Stavanger

Abstract

The Sparse Identification of Nonlinear Dynamics (SINDy) model has proven its efficacy in extracting governing equations from dynamical systems. Through sparsity-promoting techniques, SINDy offers compact and interpretable representations of dynamics. This research study demonstrates the successful application of SINDy on production data from oil wells due to good results.

Introduction

Data Driven model discovery is a field that is rapidly developing with a range of different techniques. One of these are the SINDy algorithm, also known as sparse identification of non-linear dynamical systems.

SINDy method was first introduced in 2016 by S. L. Brunton, J. L. Proctor and J. N. Kutz. This was seen as a powerful technique to identify nonlinear dynamical systems from data.

There has also been developed an open-source python package, pySINDy, to utilize the SINDy method easier. It's open source and has been developed over the last years. This is a good tool both for beginners and advanced users as well.

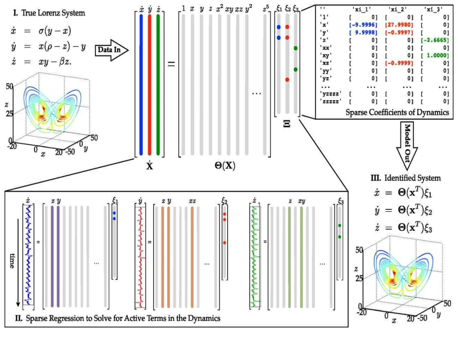
Methodology

The SINDy method uses a set of data as input and differentiate the data. Then a custom library with candidate terms will be generated. These libraries can contain d'th–degree polynomials, exponential terms, sine and cosine. It's only limited by one's imagination.

Using the data and the library of candidate functions, the SINDy algorithm searches for the sparsest set of candidate functions that can represent the dynamics observed in the data.

The result of the SINDy algorithm is a set of governing equations that best capture the system's dynamics.

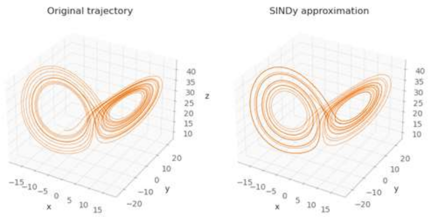
Overview of the SINDy method



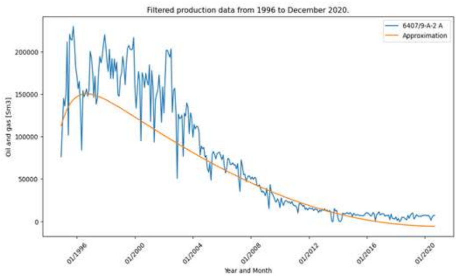
Results

The results are best shown by comparing graphs. There will be results from two different tests. One is a more theoretical result where the data is generated and are the result from the Lorenz system. The two other is based on data from Diskos from a well on the Draugen oil field and the Statfjord Øst field. These system has more noisy data compared to the Lorenz system.

Lorenz system

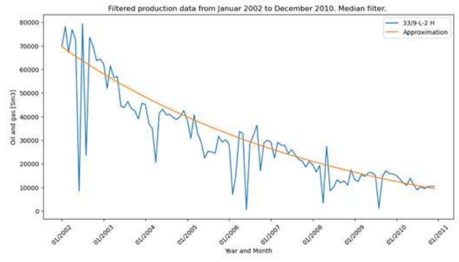


Draugen



More results

Statfjord Øst



Conclusion

The results was promising and good. There can be done a lot of testing with different parameters and changing the library with candidate terms. It shows that even if the data from Diskos was a bit spiky the SINDy method still managed to find some equations to capture the system's dynamics. Even though the results was promising the method is sensible to noise, and a bit of filtering before and during the differentiation was necessary to get good results. Good knowledge about the system is also a benefit when creating the library with candidate functions.

Acknowledgements

I would like to extend a big thank you to my supervisor Aksel Hiorth, who has been a great help during the process of my master's thesis.