Computers & Geosciences 85 (2015) 16–25

Contents lists available at ScienceDirect

Computers & Geosciences

journal homepage: www.elsevier.com/locate/cageo

Case study

Geological feature selection in reservoir modelling and history matching with Multiple Kernel Learning

V. Demyanov*, L. Backhouse, M. Christie

Institute of Petroleum Engineering, Heriot-Watt University, Edinburgh, UK

A R T I C L E   I N F O

Article history:
Received 23 December 2014
Received in revised form
21 July 2015
Accepted 22 July 2015
Available online 26 July 2015

Keywords:
Kernel learning
Uncertainty quantification
History matching
Brugge case study

A B S T R A C T

There is a continuous challenge in identifying and propagating geologically realistic features into reservoir models. Many of the contemporary geostatistical algorithms are limited by various modelling assumptions, like stationarity or Gaussianity. Another related challenge is to ensure the realistic geological features introduced into a geomodel are preserved during the model update in history matching studies, when the model properties are tuned to fit the flow response to production data. The above challenges motivate exploration and application of other statistical approaches to build and calibrate reservoir models, in particular, methods based on statistical learning.

The paper proposes a novel data driven approach – Multiple Kernel Learning (MKL) – for modelling porous property distributions in sub-surface reservoirs. Multiple Kernel Learning aims to extract relevant spatial features from spatial patterns and to combine them in a non-linear way. This ability allows to handle multiple geological scenarios, which represent different spatial scales and a range of modelling concepts/assumptions. Multiple Kernel Learning is not restricted by deterministic or statistical modelling assumptions and, therefore, is more flexible for modelling heterogeneity at different scales and integrating data and knowledge.

We demonstrate an MKL application to a problem of history matching based on a diverse prior information embedded into a range of possible geological scenarios. MKL was able to select the most influential prior geological scenarios and fuse the selected spatial features into a multi-scale property model. The MKL was applied to Brugge history matching benchmark example by calibrating the parameters of the MKL reservoir model parameters to production data. The history matching results were compared to the ones obtained from other contemporary approaches – EnKF and kernel PCA with stochastic optimisation.

1. Introduction

Sub-surface reservoir characterisation is subject to vast uncertainties that make prediction of the reservoir dynamic behaviour a challenging task, which usually involves history matching to calibrate a model to dynamic data. Uncertainties in reservoir characterisations are of different kinds: (i) data uncertainty associated with the observation/calibration errors (as depicted in Arnold et al., 2013); (ii) model uncertainty related to the model description (e.g. assumptions, equations, parameters etc., see Massonnat, 2000); (iii) model solution errors subject to the numerical solver algorithm and discretisation accuracy used (e.g. in O'Sullivan and Christie, 2005); (iv) model inadequacy representing the physics missing from the model that can be accounted for by other means (e.g. after Kennedy and O'Hagan).

History matching (HM) of reservoir models to dynamic data provides a way to infer model uncertainty to make more accurate predictions. Traditionally history matching becomes an exercise in inferring the model parameter values, that are chosen based on the given model description. Inferring uncertainty related to geological model description through history matching remains challenging, because it often needs to be done across a set of model with different parameters and even equations (e.g. Gaussian vs Boolean).

The static model or "geomodel" encapsulates the best understanding of the relationship between geological and petrophysical parameters based on the geological interpretation, e.g. depositional environment, selection and description of facies, etc. Geological interpretation is one of the main uncertainties commonly biased towards a subjective opinion and is often difficult to rigorously quantify (Bond et al., 2007). Uncertainty from the

* Corresponding author.
E-mail address: vasily.demyanov@pet.hw.ac.uk (V. Demyanov).

http://dx.doi.org/10.1016/j.cageo.2015.07.014
0098-3004/

geological interpretation may often prevail over the uncertainty coming from the chosen set of the model parameters or the subsscale variability, the latter is represented by the stochastic nature (seed) of a geomodelling algorithm. Therefore, it is important to history match across multiple plausible model descriptions/geological interpretations where relevant. Geostatistical algorithms are traditionally used to describe the distributions of spatial properties in geomodels. Geostatistical algorithms are good in establishing the appropriate level conditioning to geological measurement data. A range of stochastic geostatistical algorithms based on a covariance/variogram function, objects or training images are routinely used to populated the geomodel grid with lithological and petrophysical properties (Chiles and Delfiner, 2009; Mariethoz and Caers, 2014). All of them involve evaluation of the model dependent parameters: object geometries and placements in the object-based models; variogram correlation range and anisotropy for Gaussian or Indicator simulations; training image definition and appropriate local transformation for multipoint statistics (MPS) models.

These parameters are often inferred in a history matching task within the single chosen set of model equations, e.g. using a single training image or a certain set of object shapes (with limited allowed transformations). For example, the history matching of the variogram model parameters of a Gaussian Random Field model description was done through a Bayesian uncertainty quantification framework (Demyanov et al., 2004). History matching of channelised reservoir training image based models with realistic prior information was done in (Rojas et al., 2014a,b). More recently, the challenge in accounting uncertainty across multiple model definitions in history matching has been tackled in (Park et al., 2013; Rojas et al., 2013). They considered different geological interpretations with multiple training images and implements history matching in the distance metric space.

Calibration of geostatistical models to dynamic data still remains a challenge for the industry, though a lot of research has been done in developing history matching techniques that include gradual deformation method (GDM) (Hu, 2000); probability perturbation approach (Caers and Hoffman, 2006), which extends GDM to probabilistic Bayesian formulation; adaptation of population based stochastic optimisation algorithms (Mohamed et al., 2010a,b; Schulze-Riegert et al., 2002), other gradient based approaches for this problem (Gomez et al., 2001; Bissell et al., 1997).

Among other approaches, emerged more recently, there is a pattern based history matching with multi-point statistics facies modelling in (Melnikova et al., 2015), where spatial probabilistic information is elicited from training images and then inferred in a Bayesian way. Ensemble based data assimilation approach, such as EnKF, have been also widely used to evaluate uncertainty with an ensemble of calibrated models (Evensen, 1998). Ensemble based approaches have an advantageous capability of more flexible integration of prior knowledge – in a form of initial prior set of models, which are gradually assimilated through perturbation of the model state to fit the observed dynamic response. Such setting is suitable to account for multiple possible prior model states, which represent uncertainty in geological reservoir description (e.g. as in Peters (2008)). However, preserving geological reslim in the posterior ensemble still remains a challenge. Several history matching studies have been performed following this approach (Oliver and Chen, 2011; Chen et al., 2010; Mohamed et al., 2010a). Clustering across an ensemble of geological realisations was implemented in adaptive sparse model representation in a history matching study by Khaninezhad and Jafarpour (2014).

In this paper we propose a way to history match a reservoir model across different possible prior geological scenarios (model descriptions). We use a data driven kernel based model to populated petrophysical reservoir properties to integrate multiple types of data: observed data from wells, soft seismic information and multiple geological concepts (prior ensemble). Multiple Kernel Learning (MKL) model selects relevant spatial features from multiscale input information. The MKL model is then history matched to production data using adaptive stochastic sampling (particle swarm optimisation).

2. Modelling approaches

2.1. Data integration in reservoir modelling

Traditional geostatistics have been used for efficient data integration in reservoir modelling. Geostatistical paradigm allows integrating point and pattern data through model conditioning to data with a linear regression under a two-point covariance spatial relationship. Secondary (soft) correlated information can be integrated through a linear relation, e.g. kriging with external drift (Chiles and Delfiner, 1999) or collocated co-kriging (Almeida and Journel, 1994), see Annexure for more details.

More recently multipoint statistical moments have been implemented in geostatistical prediction (Strebelle, 2002; Dimitrakopoulos et al., 2010) to extend the flexibility of spatial correlation model beyond the two-point covariance. Multi-point statistics approach is superior to a classical two-point statistics, because it provides a richer, more flexible and a generalised description of spatial correlation (Mariethoz and Caers, 2014), see Annexure for more details. However, it is still subject to the stationarity assumption of some form (training image) and it has a limited linear capacity in integrating multivariate and multi-scale secondary information (Strebelle and Zhang, 2005; Hu and Chugunova, 2008). The problem of non-stationarity is often addressed through accounting for a trend, which usually implies a linear relation with the simulated probability field. A more recent pattern simulation algorithm, which handles non-stationary training images using a distance metric approach, was proposed in (Honarkhah and Caers, 2012).

In the present work we will address the problem of improving the flexibility of spatial reservoir property model description with a non-linear integration of spatial information from multiple patterns, which represent possible geological concepts. This approach provides more flexibility in data conditioning within a statistical learning paradigm and is not restricted by stationary assumptions, since it is purely data driven.

2.2. Statistical learning in reservoir modelling

In recent years statistical learning (Vapnik, 1995) has become more intensively used for modelling geological reservoir properties. Data driven algorithms have shown to be an efficient alternative to traditional modelling approaches for difficult problems, e.g. dealing with noisy data or non-stationary cases. Kernel learning methods – Kernel PCA, Support Vector Machines (SVM), Multiple Kernel Learning (MKL) – have been applied for modelling and classification of reservoir properties (Sarma et al. 2008; Demyanov et al., 2008, 2011; Al-Anazi and Gates, 2010). Kernel based methods have been also used for multi-dimensional scaling of model realisations into the metric space for more representative ranking and description of relations between different geological scenarios (Scheidt and Caers, 2009).

In earlier work (Demyanov et al., 2008) we have demonstrated capabilities of the semi-supervised support vector regression (SVR) to successfully model petrophysical property distributions in a fluvial reservoir. SVR model computes linear regression in high dimensional space (the reproducing kernel Hilbert space) using a single kernel for implicit mapping of the data. However, a single