Heriot-Watt University
Research Gateway

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

Citation for published version:
Demyanov, V, Backhouse, LJ & Christie, MA 2015, 'Geological feature selection in reservoir modelling and history matching with Multiple Kernel Learning', *Computers and Geosciences*, vol. 85, no. Part B, pp. 16-25. https://doi.org/10.1016/j.cageo.2015.07.014

Digital Object Identifier (DOI):
10.1016/j.cageo.2015.07.014

Link:
Link to publication record in Heriot-Watt Research Portal

Document Version:
Peer reviewed version

Published In:
Computers and Geosciences

Publisher Rights Statement:
Copyright © 2015 Elsevier Ltd.

General rights
Copyright for the publications made accessible via Heriot-Watt Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
Heriot-Watt University has made every reasonable effort to ensure that the content in Heriot-Watt Research Portal complies with UK legislation. If you believe that the public display of this file breaches copyright please contact open.access@hw.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 18. May. 2025

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.

vasily.demyanov@pet.hw.ac.uk

1 Abstract

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.

Keywords: kernel learning, uncertainty quantification, history matching, Brugge case study

2 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; (ii) model uncertainty related to the model description (e.g. assumptions, equations, parameters etc.); (iii) model solution errors subject to the numerical solver algorithm and discretisation accuracy used (e.g. in O' Sullivan & Christie, 2005); (iv) model inadequacy representing the physics missing from the model that can be accounted for by other means (e.g. after Kennedy & 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 geological interpretation may often prevail over the uncertainty coming from the chosen set of the model parameters or the subs-scale 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 J-P. and Delfiner P. 2009; Mariethoz G., Caers J., 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., 2014 a&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, Hoffman, 2006), which extends GDM to probabilistic Bayesian formulation; adaptation of population based stochastic optimisation algorithms (Mohamed et al., 2010, Schulze-Riegert et al., 2002), other gradient based approaches for this problem (Gomez et al., 2001; Bissel 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 et al, 2011; Chen et al 2010; Mohamed et al, 2010 a). Clustering across an ensemble of geological realisations was implemented in adaptive sparse model representation in a history matching study by Khaninezhad and Jafapour, 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 multi-scale input information. The MKL model is then history matched to production data using adaptive stochastic sampling (particle swarm optimisation).

3 Modelling Approaches

3.1 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 & Journel, 1994), see Annex for more details.

More recently multipoint statistical moments have been implemented in geostatistical prediction (Strebelle, 2002; Dimitrakopoulos, 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 G., Caers J., 2014), see Annex 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 et al., 2005; Hu et al, 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 & 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.

3.2 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; Alanazi 2009). 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 & 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 kernel makes it difficult to reproduce multi-scale non-stationary spatial patterns. The difficulty was partly overcome by using semi-supervised learning approach with unlabelled data, which, however, does not explicitly reflect the correlation on multiple scales. The Multiple Kernel Learning (MKL) model applied in this paper is a generalisation of SVR that allows multiple kernel representation for every separate model input. For instance, when modelling a gridded porosity distribution conditioned to core data from wells and seismic, separate kernels will be used for spatial co-ordinates and the soft multi-component seismic input to account for information at different scales (well scale, seismic components). Thus, MKL provides capability to rigorously integrate different multi-scale information, which represents various geologically relevant features. For instance, MKL was applied for modelling spatial wind fields in Foresti, et al. 2011. Demyanov et al. (2011) used MKL to select relevant spatial scales derived from seismic in a history matching study.

In the present work we extend MKL application in reservoir modelling in attempt to integrate spatial information from multiple possible geological scenarios. MKL based geomodel framed in a history matching context aims to select spatial features from those geological scenarios that are more likely to support production data.

3.3 Feature selection in History Matching Workflow with Multiple Kernel Learning

Feature (or variable) selection is the technique whereby a subset of variables is chosen from the available set. This subset includes those variables most relevant in predicting the response or outcome. Feature selection applied for geomodelling allows extracting the most relevant information from the input data and creating a multitude of possible geological realisations previously not available.

One of the principle problems in reservoir modelling is to describe the relations between the observed spatial information and the target predicted variables – porosity and permeability. These relations are

highly complex and subject to many uncertainties. The group of all possible geological scenarios along with all other available information, such as seismic, outcrop, etc., are known as variables used in reservoir description. The problem is to which extent a model needs to be conditioned to each piece of data and how to relate different bits of the data within the model in the most relevant and geologically plausible way. We have already mentioned that among the limitations of geostatistics are hard kriging conditioning under stationarity and linear relations with soft conditioning information. MKL, as a data driven approach, is capable of finding dependencies embedded in the input geological features and integrating them into the spatial property prediction model through a non-linear weighting. For instance, in case of seismic, which is commonly used as soft conditioning data in geostatistical modelling, the challenge is to determine the influence seismic data have on the prediction of porous properties.

MKL shows distinct advantages over competing methodologies such as Kernel PCA and EnKF. The weights assigned to each feature in MKL are interpretable in that they sum to one. From this one can ascertain those variables that are the most important in the prediction as well as those that are redundant. The weight is applicable to each input variable or feature and not to some mapped or transformed variable as it is in Kernel PCA. A further disadvantage of KPCA is that the output suffers from the supposed “pre-image” problem. That is the output has to be inversed mapped into the input space that is usually far larger than the mode response output space due to the narure of the inverse problem (Bakır et al., 2004).

Feature selection is carried out by means of an optimisation through a training process (see Figure 1). MKL computes prediction of the target variable with the selected features, unlike other machine learning algorithms (e.g. artificial neural networks), where the feature selection (choice of inputs) is a pre-conditioning step applied prior to training the algorithm. Tsamardinos et. al (2003) pointed out that the selection of the relevant features may well be a subset rather than individual features. That is, a group of features are selected together to produce the MKL prediction. Different combinations of selected features identify a range of model description scenarios driven by data rather than by pre-defined fixed assumptions.

MKL has the advantage of being efficient in that feature selection is not a process carried out either before or after the training. MKL is efficient as a feature selection technique in that relevant features are selected automatically in the tuning of the parameters stage. This is unlike other methods where feature selection is performed first then the estimation model is run.

In this paper we will use a dual loop to tune MKL reservoir model for porosity (φ) and permeability (k): (i) the inner loop optimises the impact of the input features by minimising the mismatch of the static model; (ii) the outer loop tunes MKL hyper-parameters to minimise the dynamic mismatch of the reservoir model response (see Figure 1). We will also discuss the differentiation in the impact of the input features, which represent prior geological scenarios.

FIGURE 1 HERE

MKL has been used as a feature selection technique in various diverse applications: Tuia. D., et. al. (2010) applied MKL to select relevant features from satellite images; Takashima et al. (2011) applied MKL to select the relevant spectral dimension of the acoustic transfer function in sound source localization; Dileep et. al. (2009) details the theory of MKL feature selection and applied it to an image classification problem.

The data driven MKL predictor is still subject to the choice of the parameter values, which define the kernels for each input feature (σ_i), the complexity bound C on the support vectors and the error margin ε (for more details on the support vector regression refer to section 4.1 and Figure 3). Ad-hoc trial and error choice of these most influential parameters may be quite risky. He et al. (2008) used a genetic algorithm to tune the parameters of their feature selection algorithm MO-SVR iteratively analysing the scale of each factor (σ of the Gaussian Kernel). Kanevski et al. 2009 and Foresti, 2011 followed a cross-validation/testing approach in picking optimal combination of kernel parameters from the test error surface.

In this work we use a history matching loop to tune the MKL model parameters (σ_i, C, ε) using particle swarm optimisation (PSO, Kennedy & Eberhart, 1995) (see Figure 1). This includes populating reservoir grid with porosity (φ) and permeability (k) and the subsequent flow simulation to compute the

206 dynamic model response. MKL model parameters are updated iteratively guided by the mismatch
207 between the history and the dynamic model response. The MKL prediction model is then re-trained to
208 update the feature selection (through the weights w_i) with every update in its parameters.

209 ## 4 Theory of the algorithms

210 This section overviews kernel learning algorithms applied for porosity/permeability field prediction.
211 First, we briefly overview support vector regression, for which MKL is a generelisation.

212 ### 4.1 Support Vector Regression (SVR)

213 Support vector regression (SVR) is a function estimation extension of the well-known support vector
214 machine classifier developed within the framework of the statistical learning theory (Vapnik, 1995). By
215 controlling model complexity SVR provides robust solutions to non-linear and high-dimensional
216 regression problems.

217 SVR belongs to the family of kernel methods that allow finding linear solutions to non-linear regression
218 problems by applying an implicit Hilbert space mapping to original input data. Such mapping is carried
219 out by means of kernel functions. Kernels can be used to replace dot products between training points in
220 a higher dimensional (Hilbert) space, where the original data are mapped. More formally, the so called
221 "kernel trick" can be defined as

222 $$K(\boldsymbol{x}, \boldsymbol{x}_i) = \psi(\boldsymbol{x}) \cdot \psi(\boldsymbol{x}_i) \tag{1}$$

223 where ψ denotes the mapping of an input vector $\boldsymbol{x}$ to the Hilbert space. Among the broad list of
224 kernels which have been tested in different studies (Kanevski et al. 2009), we have chosen a Gaussian
225 RBF kernel for this paper.

226 SVR prediction at $\boldsymbol{x}$ becomes as a weighted expansion of kernels (Schölkopf and Smola, 2002):

227 $$f(\vec{x}) = \sum_{i=1}^{N_{SV}} \alpha_i K(\boldsymbol{x}, \boldsymbol{x}_i; \theta) + b \tag{2}$$

where α_i are Lagrangean multipliers, $K(\boldsymbol{x},\boldsymbol{x}_i;\ \theta)$ is the kernel function with its internal hyper-parameters θ, b is the bias and N_{sv} is the number of support vectors.

A key component of the SVR model formulation is the definition of an a-priori level of noise in the data as can be seen in Figure 2. Such robust cost function, also called "ε-tube" defined by the hyper parameter ε, allows building sparse models using only a reduced set of the data (called support vectors). The amount of noise assumed is controlled by the parameter ε. The amount of penalisation for the data lying outside the ε-tube is controlled by the parameter C, which is the bound for the support vectors. The data points which lie outside the ε-tube are penalised according to a free complexity parameter C yielding a regularized solution.

FIGURE 2 HERE

4.2 Multiple Kernel Learning (MKL)

MKL predictor can be written in a non-linear kernel form:

$$Z(x) = \sum_{m=1}^{P} w_m \sum_{i=1}^{N_{SV}} \alpha_i K_m(\boldsymbol{x}, \boldsymbol{x}_i, \sigma_m, C, \varepsilon) \tag{3}$$

where σ_m denotes the internal parameter of the single kernel and w_m its respective weight.

Thus, a multiple kernel forms a weighted convex combination of m basis kernels. The regression eq. 3 is similar in some way to the kriging equation (see Annex eq. A.1), further more w_m sum to 1, but are also positive definite (Lanckriet et al., 2004). The principle difference is that the linear kriging regression weighs the outputs – the observed values of the target variable $Z(\boldsymbol{x}_i)$; while MKL is a regression of the inputs in the kernel form $K(\boldsymbol{x}, \boldsymbol{x}_i)$. This is provided through implicit mapping of data points from the input space to a reproducing kernel Hilbert space (RKHS). The resulting space induced by the multiple kernel is the sum of all RKHS for which the weights w_m's are non-zero (feature selection). The kernels introduce a non-linear element and correspond to the support vectors α_i, which are determined for each input feature through training. Impact of the relevant features is controlled by the weights w_m, i.e. the selected features get higher weighting. The optimal weighted combination can

give insights about the relevance of the input features/kernels. The multiple kernel formulation brings a lot of flexibility by handling a variety of input features and kernel hyper-parameter tuning. A multi-scale processes can be modelled by using two or more kernels of different bandwidths σ_m acting on the same input feature (Pozdnoukhov et al., 2008). Data integration can be approached by considering one kernel for each source of information (e.g. Demyanov et al., 2011).

Several algorithms have been proposed to solve the MKL problem efficiently, i.e. to find the optimal weighted combination (Lanckriet, et al., 2004; Sonnenburg, et al., 2006). In this paper we used SimpleMKL algorithm proposed by Rakotomamonjy et al., 2008, where the convex problem is solved with a dual loop optimisation through gradient with line search and descent direction update (implemented in the inner loop in Fig. 1). In this work we have applied a Simple MKL implemented in the similar way as in (e.g. Demyanov et al., 2011):

Table 1. MKL Algorithm

initialise MKL parameters *C* and σ for each kernel for both porosity and permeability models within prior ranges

repeat

- train MKL with the given parameters and predict porosity (Feature selection loop in Fig. 1)
- add the predicted porosity to the chosen input features, re-train MKL to predict permeability
- run flow simulation on the reservoir reservoir grid populated with porosity and permeability and forecast oil production from the
- compute misfit w.r.t. to historical oil production data for each of the 20 production wells
- update *C*, ε and kernel widths σ by adaptive stochastic sampling from the prior

until convergence of parameters and matching of oil/water production (observed at 500 iterations)

5 History Matching with MKL Reservoir Model

5.1 Brugge Reservoir Case Study Description

Feature selection problem was formulated for Brugge reservoir model. Brugge case study was designed as a benchmark problem for history matching and optimisation (Geel, 2008; Peters et al., 2010). The reservoir consists of four stratigraphic zones: fluvial (deltaic), lower shore face, upper shore face and sandy shelf (Geel, 2008). The unique characteristic of this model is the provided prior ensemble of grids

of the porous properties, which represent the spread of geological uncertainty according to the initial beliefs about the reservoir geology. The prior ensemble of 104 unconditional realisations designed by TNO includes the models that are described by different geostatistical algorithms (objects based and variogram based) and with variable assumptions about the facies and properties correlations (Geel, 2008). Prior realisations (Fig. 3, sand shelf layer shown only) do not match the production data, which sets up a history matching challenge.

FIGURE 3 HERE

Stratigraphy of Brugge reservoir includes four zones (split into 9 grid layers), which represent a typical sequence of depositional environments in a Brent reservoir: fluvial, lower shoreface, upper shoreface and a sandy shelf. Each layer features its own spatial structure, correlation of porosity and permeability and their statistical distribution properties.

5.2 MKL reservoir property model

Feature selection with MKL aims to select the most relevant geological features from the ensemble of prior realisations based on how well the combination of these features fit the observed well data and match the production history. MKL is capable of non-linear blending of the selected spatial features to provide the right balance of each bit of the input data (from the wells) and weigh their impact on the property distribution at a distance.

MKL was applied to predict porosity (φ) and permeability (k) fields from the input features following the workflow in Fig. 1 and the training algorithm (see Table 1). MKL was trained using the target (output) on porosity and permeability data from 30 well locations interpreted from the wireline logs, which slightly vary from the properties observed across the prior ensemble of 104 realisations. The gridded porosity and permeability from the prior realisations were used as the inputs for the MKL model. In case of the permeability predictor – the modelled porosity was used as an additional input to establish the petrophyscial correlation between the two properties derived from the data. Seismic data (4D) as pressure and saturation differences (see Fig. 4) were used as additional inputs. Separate MKL models were defined for each of the four stratigraphic zones corresponding to the different depositional

environments: fluvial (2 grid layers), lower shore face (3 grid layers), upped shore face (3 grid layers) and sandy shelf (1 layer). This makes altogether 8 MKL predictors for φ and k. Kernel transformation of the input features in to porosity and permeability fields by the MKL predictor is schematically illustrated in Fig. 5. Figure 6 illustrates MKL porosity models computed with different hyper-parameter values.

FIGURE 4 HERE

FIGURE 5 HERE

FIGURE 6 HERE

5.3 History Matching Model set-up

The history matching challenge has been tackled in many studies using different approaches (Peters et al., 2010). Although, the Brugge benchmark was very suitably designed for EnKF application, a comparative study (Mohamed et al., 2010b) found that history matching with Particle Swarm optimisation (PSO) was able to provide matches of equally good quality as the EnKF ones but different between each other.

History matching of Brugge model often requires re-parametrisation of the geological property distribution to introduce a common metric for the prior realisations, for instance kernel PCA was implemented in (Mohamed et al, 2010 b). This can be seen as a potential weakness as the realisations evolved through history matching may lose geological realism and consistency with the prior geological assumptions. Keeping the updated model realisations within the geological prior is not a straightforward problem once it comes to patterns and uncertainty in geological model description. An approach for eliciting prior information from patterns was proposed in (K. Lange, et al. , 2012), where the prior pdf of patterns elicited from a training images was inferred. This is a plausible approach, however, questions still remain: how much information is there in the a priori training image to elicit and what are the means by which it should be elicited in the most effective way.

This paper attempts to find some answers through the kernel learning route. So, it is for the data to prompt how much and what king of information is to be inferred from the prior into the posterior. We have sampled the MKL parameters from the prior. We have checked that the generated models are consistent with the spatial correlation structures reflected in the prior patterns (which is some representation of the a priori). MKL hyper-parameters require tuning, which was done through history matching to calibrate the reservoir model flow response to production data. The historical data from the past 10 years were used in to calibrate the MKL property model in an inverse loop (similar to Demyanov et.al. 2011, where a Neighbourhood approximation was used). The least square misfit between the flow simulation response and the historical production (pressure and rates at the individual wells) ranked the match quality and drove the history matching process (after Mohamed et al. 2010b).

In this work we used Particle Swarm Optimisation algorithm (PSO, after Kennedy & Eberhart 1995) sample iteratively from the prior uniform ranges of the MKL hyper-parameters parameters: kernel bandwidth for each input feature σ_i, regularisation constant C and error margin ε (see the outer loop in Fig. 1). The prior intervals for the kernel bandwidths σ_i were defined as a fraction of the corresponding feature range. The prior interval for the regularisation parameter C, which controls the roughness of the predictions, was chosen according to the uncertainty of the maximum value for the predicted property.

MKL predictions of porosity and permeability reservoir grids computed for every combination of the kernel parameters were then the flow simulated (see Fig. 1). The evolution of the MKL kernel parameters through history matching iterations is illustrated in Fig. 7, where the models with lower misfit tend to home in around the particular bandwidth values corresponding to the better history match models. These clusters of models are different for the four reservoir zones of the Brugge reservoir.

FIGURE 7 HERE

Sampling for the optimal MKL parameters has resulted in a fairly good history match on both field level and some of the individual wells (see Fig. 8 a,b). The best match was obtained after 2000 iterations and a number of low misfit models were generated for consecutive iterations (Fig. 8c). Multiple history matches obtained as a result of the stochastic PSO algorithm provide a spread of

uncertainty in the production response. PSO set up was chosen to balance the exploration and exploitation components in HM model search according to earlier studies (Mohammed 2010a). There were 20 particles with cognitive and social component multipliers equal to 1.333 and inertia decay rate of 0.9. Each HM run was taken through 3000 iterations that included an MKL call and a flow simulation each. The CPU time for one 3000 iterations run was 45 h on a 10 node Xeon cluster.

The lowest misfit for the best history match achieved was around 180,000, which is better than the misfit of around 300,000 thousands obtained with kernel PCA model in (Mohamed et al, 2010b). Figure 9 compares the quality of the misfit in the individual wells with the earlier results from Mohamed et al, 2010b). It is clear that various methods result in comparative quality match, however, the match quality obtained by one or another method differs from one well to another.

The use of 4D seismic data (saturation and pressure differences) as two additional MKL inputs improved was quite important the porosity & permeability predictors. It resulted in taking the history match misfit down by 50% vs using just the 104 prior geological realisations as the input. The weight of the seismic features was a few %, which was comparable to the weight of the selected features, although would vary from one model to another.

FIGURE 8 HERE

FIGURE 9 HERE

6 Results of MKL Feature Selection

Feature selection with MKL demonstrated that not all out of the 104 input realisations contributed equally to the property prediction. The MKL weights w_m from eq. (3) determine the contribution of each input feature x. There were, usually, no more than a dozen of input features that contribute in total over 90% of the weighting. An example distribution of the weights with the corresponding input realisations is illustrated in Fig. 10.

FIGURE 10 HERE

FIGURE 11 HERE

Further investigation of the selected input features revealed that they correspond to different modelling algorithms and facies assumptions. It appeared that each of the selected features represent a particular spatial scale from 500 m to 4000 m (see Fig. 11) and their weighting demonstrates a fair spread across the scale coverage. This leads to an important conclusion about MKL being capable to select multi-scale information from patterns and then to blend it together in the prediction. Spatial correlation analysis using variorgams illustrates quite complex and multi-scale character of the spatial geological patterns selected to be blended in the MKL regression model (see Fig. 12). Three out of the seven selected realisations feature anisotropic spatial correlation with the range (geometric) anisotropy. The total weight of the anisotropic input realisations contributes to over 30% (see Fig. 12) with the rest selected inputs being isotropic.

FIGURE 12 HERE

MKL reservoir model calibration in stochastic history matching framework results in multiple good fitting solution. These solutions are non-unique due to the stochastic nature of the history matching algorithm (PSO). Therefore, different combinations of input features (and consequently the kernel algorithm parameters) provide similar quality matches. It was noticed that some of the input realisations had been repeatedly selected in multiple history matching runs. However, their weights determined through the MKL training process vary from one history match model to another. Figure 14 illustrates the variation of the weighting factors for the three most often selected input realisations across 20 history matching runs. Solutions from multiple history matching runs provide the robustness of the uncertainty spread in the production predictions. **FIGURE 13 HERE**

7 Conclusions

The paper demonstrates an application of multiple kernel learning (MKL) to the problem of selecting of relevant geological features to describe geological uncertainty in a reservoir history matching study. MKL proved to be an efficient tool for modelling complex spatial distributions of reservoir properties

(porosity and permeability) with a non-linear and multi-scale character. MKL can be seen complimentary and even advantageous to certain extent to traditional geostatistical algorithms for its flexibility in integration of diverse spatial multivariate information. However, geostatistical tools are still needed for modelling geologically realistic input features for MKL. MKL provides a rigorous balance between fitting a kernel regression model to data, which are subject to noise and possible interpretational uncertainty, and the model complexity. Model complexity is controlled with the support vector formalism described by statistical learning theory (Vapnik, 1995).

MKL reservoir model is applied to a history matching problem in Brugge case study. Geological uncertainty is described by a set of prior realisations, which represent a range of different geological model descriptions based on the algorithmic and the model assumption choices. MKL does not assume any particular geological description but rather selects relevant spatial patterns based on the available data through training.

MKL reservoir model history matched with PSO resulted in a number models fitted to the production history. The obtained match quality is comparable or superior to the ones from the previous studies, where different model parameterisations were used with EnKF or Kernel PCA (Mohamed, 2008b). History matching with MKL demonstrated that relevant geological patterns were selected from the ensemble of prior realisations and then blended with the kernel regression in RKHS space to predict target petrophysical properties. The selected features (geological realisations) represent different spatial scales and pattern anisotropy. The resulting blends of these patterns provide solutions that match production data.

8 Acknowledgment

Funding for this work was provided by the industrial sponsors of the Heriot-Watt Uncertainty JIP (Phase IV). Brugge case study data was supplied by TNO. Foundation Herbette provided funding to visit University of Lausanne in conjunction to finalising this work.

Authors thank Mikhail Kanevski (University of Lausanne) and Dan Arnold (Heriot-Watt) for fruitful discussions and comments on the work, and Loris Foresti (University of Lausanne) for providing the

MKL code. Authors also appreciate the comments from the two anonymous reviewers that helped to improve the paper.

9 References

1. Al-Anazi, A. F., Gates I. D. 2010. "Support vector regression for porosity prediction in a heterogeneous reservoir: A comparative study." Computers & Geosciences 36.12: 1494-1503.

2. Almeida AS, Journel AG. 1994. “Joint simulation of multiple variables with a Markov-type coregionalization model”. Math Geology, 26:565–588.

3. Bakır, G., Weston, J., & Schölkopf, B. 2004. “Learning to find pre-images”. Advances in Neural Information Processing Systems, 16, 449-456.

4. Bissell, R. C., Dubrule, O., Lamy, P., Swaby, P., & Lepine, O. 1997. “Combining geostatistical modelling with gradient information for history matching: the pilot point method”. In SPE Annual Technical Conference and Exhibition. Society of Petroleum Engineers.

5. Bond, C. E., Gibbs A.D., Shipton, Z.K., Jones, S. (2007). “What do you think this is? ``Conceptual uncertainty" in geoscience interpretation”. GSA today 17.11: 4.

6. Caers, J. and Hoffman, T. 2006. “The probability perturbation method: a new look at Bayesian inverse modelling”, Mathematical geology, 38(1), 81-100.

7. Chen C., Wang Y., G Li, Reynolds AC. 2010. “Closed-loop reservoir management on the Brugge test case”, Computational Geosciences, Volume 14, Issue 4, pp 691-703

8. Chiles J-P. and Delfiner P. 2009. “Geostatistics: modeling spatial uncertainty”. Vol. 497. John Wiley & Sons.

9. Demyanov, V., Pozdnoukhov, A., Kanevski, M. and Christie, M. 2008. “Geomodelling of a Fluvial System with Semi-Supervised Support Vector Regression”, Proceedings of the VII International Geostatistics Congress, GECAMIN, Chile, pp. 627-636.

10. Demyanov V., Foresti L., Christie M. and Kanevski M. 2011. “Reservoir modelling with feature selection: a kernel learning approach”. In Society of Petroleum Engineers Reservoir Simulation Symposium, 141510-MS, pp. 503-514.

11. Dimitrakopoulos R., Mustapha H., Gloaguen E. 2010. “High-order Statistics of Spatial Random Fields: Exploring Spatial Cumulants for Modeling Complex Non-Gaussian and Non-linear Phenomena”, Math Geosciences, 42: 65–99.

12. Dileep, A. D. and Sekhar, C.C. 2009. “Representation and feature selection using multiple kernel learning”. International Joint Conference on Neural Networks. IJCNN 2009. IEEE.

13. Foresti L. 2011. “Kernel-based mapping of meteorological fields in complex orography”, PhD thesis, University of Lausanne.

14. Foresti, L., Tuia, D., Kanevski, M. and Pozdnoukhov, A. 2011. “Learning wind fields with multiple kernels”. Stochastic Environmental Research and Risk Assessment, 25:51-66.

15. Geel, K. 2008. “Description of the Brugge field and property realisations, white paper for the Comparative Case Study on Closed-loop Reservoir Management”, SPE ATW on Closed-loop Reservoir Management, Bruges, Belgium.

16. Bakır, G.H., Weston, J. and Scholkopf, B. 2004. “Learning to Find Pre-Images”, NIPS papers.

17. Gomez, S., Gosselin, O., & Barker, J. W. 2001. “Gradient-based history matching with a global optimization method”. SPE Journal, 6(02), 200-208.

18. Honarkhah, M., & Caers, J. 2012. “Direct Pattern-Based Simulation of Non-stationary Geostatistical Models”, Mathematical Geosciences, 44(6), 651-672.

19. Hu, L.Y. 2000. “Gradual deformation and iterative calibration of Gaussian-related stochastic models”, Mathematical Geology, 32(1), 87-108.

20. Hu, L. Y., & Chugunova, T. 2008. “Multiple-point geostatistics for modeling subsurface heterogeneity: A comprehensive review”. Water Resources Research, 44(11).

21. Kanevski, M., Pozdnoukhov, A., Timonin, V. 2009. “Machine Learning for Spatial Environmental Data: theory, applications and software”, EFPL Press.

22. Khaninezhad M.M., Jafarpour B. 2014. “Prior model identification during subsurface flow data integration with adaptive sparse representation techniques”, Computational Geosciences, 18 (1), pp 3-16

23. Kennedy, J., and Eberhart, R. 1995. “Particle swarm optimization”, In: Proceedings of the IEEE International Conference on Neural Networks, IEEE Service Centre, 4, 1942–1948,

24. Lanckriet, G. R. G.; Cristianini, N.; Bartlett, P. L.; Ghaoui, L. E.; and Jordan, M. I. 2004. “Learning the kernel matrix with semidefinite programming”, Journal of Machine Learning Research 5:27–72.

25. Lange K., Frydendall J., Cordua K.S., Hansen T.M, Melnikova Y., Mosegaard K. 2012. “A Frequency Matching Method: Solving Inverse Problems by Use of Geologically Realistic Prior Information”, Mathematical Geosciences, 44:783–803.

26. Mariethoz G, Caers J 2014 . “Multiple-point Geostatistics: Stochastic Modeling with Training Images”, Wiley.

27. Matheron, Georges 1969. "Part 1 of Cahiers du Centre de morphologie mathématique de Fontainebleau". Le krigeage universel. École nationale supérieure des mines de Paris.

28. Massonnat, G.J. 2000. “Can we sample the complete Geological Uncertainty Space in Reservoir Modelling Uncertainty Estimates?” SPE 38746, SPE Journal, 5 (1): 46-59.

29. Melnikova, Y., Zunino, A., Lange, K., Cordua, K.S., Mosegaard, K. 2015. “History Matching Through a Smooth Formulation of Multiple-Point Statistics”, Mathematical Geosciences, 47(4), 397-416.

30. Mohamed, L., Christie, M., Demyanov, V. 2010 a. “Comparison of stochastic sampling algorithms for uncertainty quantification”. SPE Journal 15 (01), 31-38.

31. Mohamed, L., Christie, M., Demyanov, V., Robert, E., and Kachuma, D. 2010 b. “Application of particle swarms for history matching in the Brugge reservoir”. In: Proceedings of the SPE Annual Technical Conference and Exhibition (ATCE), SPE 135264.

32. Oliver, D.S., Chen, Y. 2011. “Recent progress on reservoir history matching: a review, Computational Geosciences, 15: 185-221.

33. Park, H., Celine, C., Fenwick, D., Boucher, A. and Caers, J. (2013). “History matching and uncertainty quantification of facies models with multiple geological interpretation”, Computational Geosciences, 17 (4), 609-621.

34. Peters, E., Arts, R., Brouwer, G., and Geel, C. 2010. “Results of the Brugge benchmark study for flooding optimisation and history matching”. SPE Reservoir Evaluation & Engineering, SPE 119094–PA, 13(3), 391–405.

35. Pozdnoukhov, A., Kanevski, M. 2008. “Multi-Scale Support vector Regression for hot spot detection and modeling”, Stochastic Environmental Research and Risk Assessment, 22, 647–660.

36. Rakotomamonjy, A.; Bach, F. R.; Canu, S.; and Grandvalet, Y. 2008. “SimpleMKL”. Journal of Machine Learning Research, 9:2491–2521.

37. Rojas, T.A., Demyanov, V., Christie, M., Arnold, D. 2014. “Using Sedimentological Prior Information to Control Realism in Reservoir Models”, Journal of Petroleum Technology; 66(9):38-40.

38. Rojas, T.A., Demyanov, V., Christie, M., Arnold, D. 2014. “Controlling the Sedimentological Realism of Deltaic Reservoir Models by the Use of Intelligent Sedimentological Prior Information”, First Break; 32(10):69-72.

39. Rojas, T.A., Demyanov, V., Christie, M., Arnold, D. 2013. “Learning Uncertainty from Training Images for Reservoir Predictions”, Mathematics of Planet Earth, Lecture Notes in Earth System Sciences 2014, pp 147-151.Sarma, P., Durlofsky, L.F., Aziz K. 2008. “Kernel Principal Component

527 Analysis for Efficient, Differentiable Parameterization of Multipoint Geostatistics", Mathematical
528 Geosciences, 40(1), pp. 3-32.

529 40. Scheidt C, Caers J. 2009. "Representing spatial uncertainty using distances and kernels",
530 Mathematical Geosciences, 41 (4), 397-419.

531 41. Schoelkopf, B. and Smola, A. 2002. "Learning with Kernels". MIT Press, Cambridge MA.

532 42. Schulze-Riegert, R. W., Axmann, J. K., Haase, O., Rian, D. T., & You, Y. L. 2002. "Evolutionary
533 algorithms applied to history matching of complex reservoirs". SPE Reservoir Evaluation and
534 Engineering, 5(02), 163-173.

535 43. Sonnenburg, S., R¨atsch, G., Sch¨afer, C. and Scholkopf, B. 2006. "Large scale multiple kernel
536 learning", Journal of Machine Learning Research 7:1531–1565.9:2491–2521.9:2491–2521.

537 44. Strebelle, S. 2002." Conditional Simulation of Complex Geological Structures Using Multiple-Point
538 Statistics". Mathematical Geology, 34 (1), 1-21.

539 45. Strebelle, S. and Zhang, T. 2005. "Non-stationary multiple-point geostatistical models", in
540 Geostatistics Banff 2004, edited by O. Leuanthong, and C. V. Deutsch, Springer, pp. 235–244.

541 46. Takashima, R., Tetsuya, T., Yasuo, A. 2011. "Feature selection based on multiple kernel learning
542 for single-channel sound source localization using the acoustic transfer function", Acoustics,
543 Speech and Signal Processing (ICASSP), IEEE.

544 47. Tsamardinos I., and Constantin F. A. 2003. "Towards principled feature selection: Relevancy,
545 filters and wrappers", Proceedings of the ninth international workshop on Artificial Intelligence
546 and Statistics. Morgan Kaufmann Publishers: Key West, FL, USA.

547 48. Tuia, D., Camps-Valls, G., Matasci, G., Kanevski, M. 2010. "Learning relevant image features with
548 multiple-kernel classification", IEEE Transactions on Geoscience and Remote Sensing, 48: 3780-
549 3791.

550 49. Vapnik, V. 1995. "The Nature of Statistical Learning Theory". Springer-Verlag Berlin.

10 ANNEX

Geostatistical formalism

Conventional kriging family of algorithms accommodates hard conditioning data $Z(\boldsymbol{x}_i)$ through a linear regression, which provides unbiased estimate with a minimum variance around the mean of $Z(\boldsymbol{x})$ (Matheron, 1969):

$$Z(x) = \sum_{i=1}^{N} w_i Z(\boldsymbol{x}_i) \quad \text{(A.1)}$$

where the regression weights w_i are computed based on the spatial correlation modelled with two-point statistics (covariance or variogram).

Multi-point statistics simulations generally use training images as a source of higher-order statistics spatial correlation model and provide a probabilistic estimate of a spatial random variable A_k, which can be written in a similar linear weighting form:

$$P(A_k \,|\, Z(\boldsymbol{x}_i)) = E\{A_k\} + w \cdot (1 - E\{D\}) \quad \text{(A.2)}$$

were w is the weight derived based on the probability of A_k according to the statistics of the data events (D) distribution from the training image (multi-point statistics) (Marithoz & Caers, 2014). This generalised formulation becomes of kriging regression type (A.1) in a special case with two-point statistics.

11 LIST OF FIGURES

575 Figure 4. Pressure (a) and oil saturation (b) differences obtained from 4D seismic.

576 Figure 5. MKL porosity predictor with 106 inputs – 104 prior geological realisations and 2 seismic
577 attributes (4D pressure and saturation differences)

578 Figure 6. MKL porosity prediction maps computed with different hyper-parameter values (sandy shelf
579 layer)

580 Figure 7. Evolution of MKL tuning parameters – kernel width σ – for different reservoir zones: (a)
581 fluvial, (b) lower shore face, (c) upper shore face, (d) sandy shelf.

582 Figure 8. History matches of MKL model to production data (♦ dots) with PSO: field oil production
583 rate (a); individual well (Pro-15) production rate (b); misfit evolution through history matching with
584 PSO (c).

585 Figure 9. Comparison of the history match quality at individual wells between three methods: a) MKL
586 model with PSO HM (this paper); b) kernel PCA model with PSO HM and c) EnKF history matching
587 (after Mohamed et al., 2010 b), – green circles label good match, orange circles label fair match, red
588 circles label poor match.

589 Figure 10. Seven selected input features for the fluvial layer (prior realisations, middle) by MKL and
590 their corresponding weights (bar chart above) and the resulting MKL porosity prediction (bottom)

591 Figure 11. Spatial scale represented by each of the seven selected features for the fluvial layer (prior
592 realisations) with the corresponding MKL weights (bottom)

593 Figure 12. Variogram analysis of the selected input features: omnidirectional variograms for the seven
594 selected input features represent different spatial correlation scale; bottom right – spatial anisotropy
595 (geometrical) direction according to the variogram ranges of the three features (from the selected seven)
596 with the corresponding MKL weights, spatial correlation for the rest four selected realisations is
597 isotropic

598 Figure 13. Variation of the contribution of a selected input features across multiple HM models
599 obtained from 20 runs: three features, which represent different spatial scales, were selected in each of
600 the 20 HM matches, but the weight of their contribution varies from one HM to another.

601

Figure(s)

Click here to download high resolution image

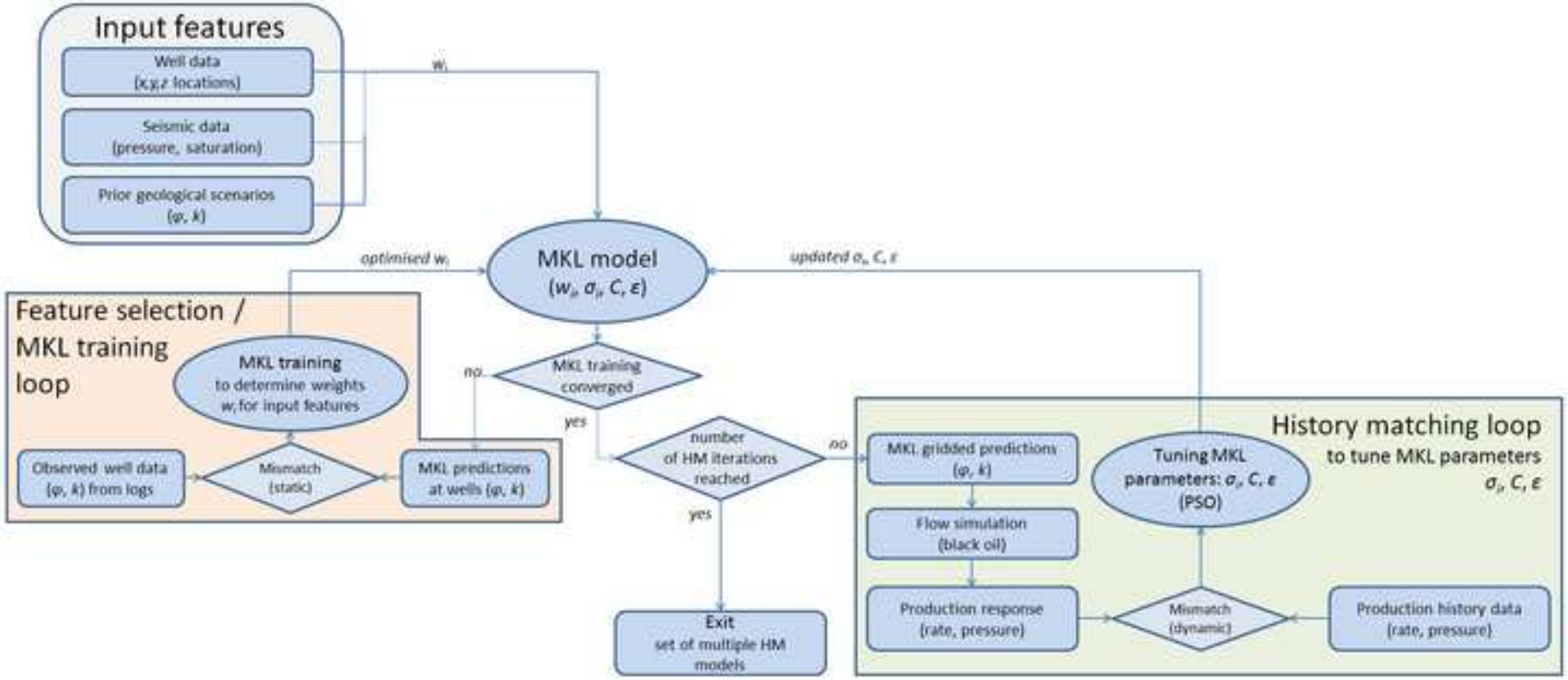

Figure 2

Click here to download high resolution image

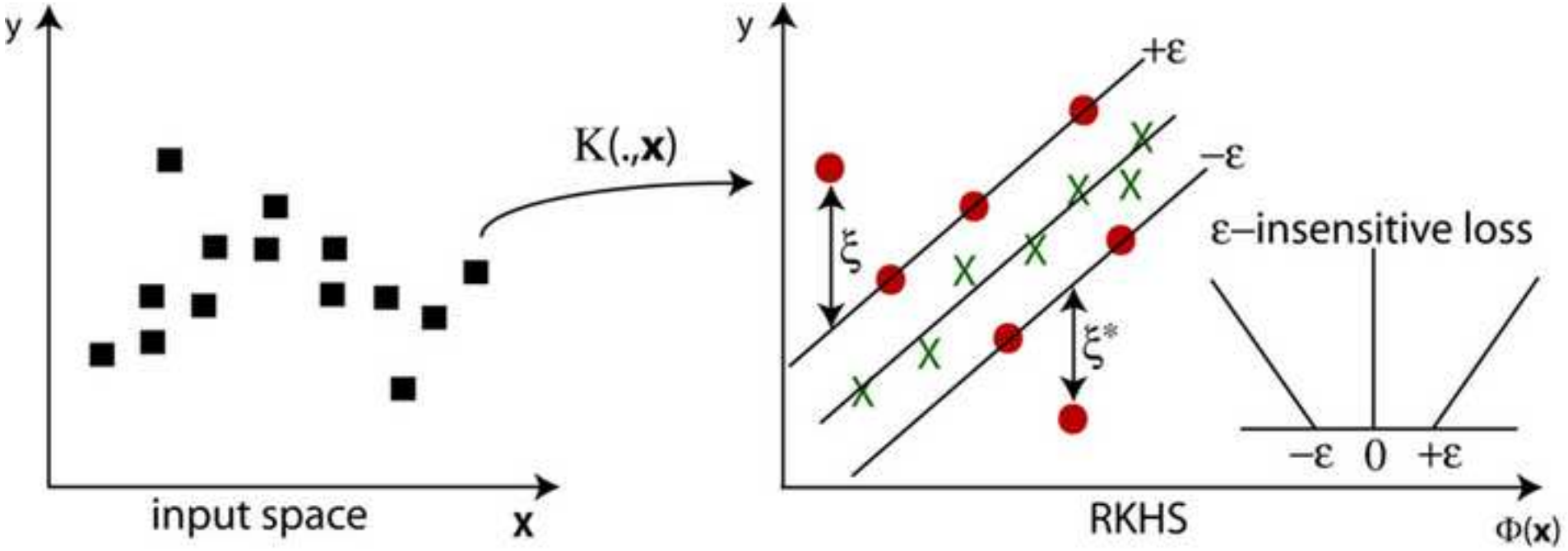

Figure 3

Click here to download high resolution image

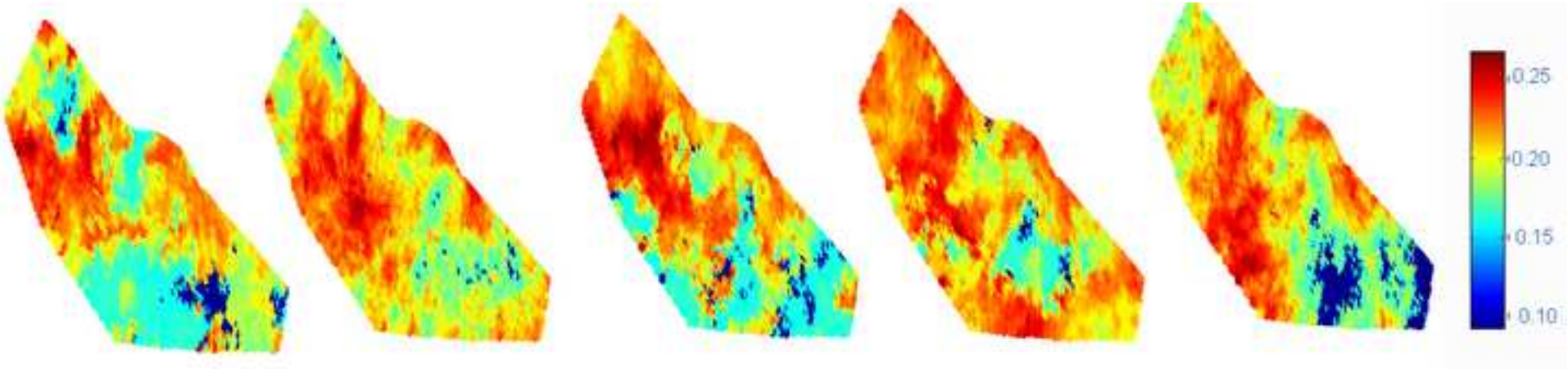

Figure 4

Click here to download high resolution image

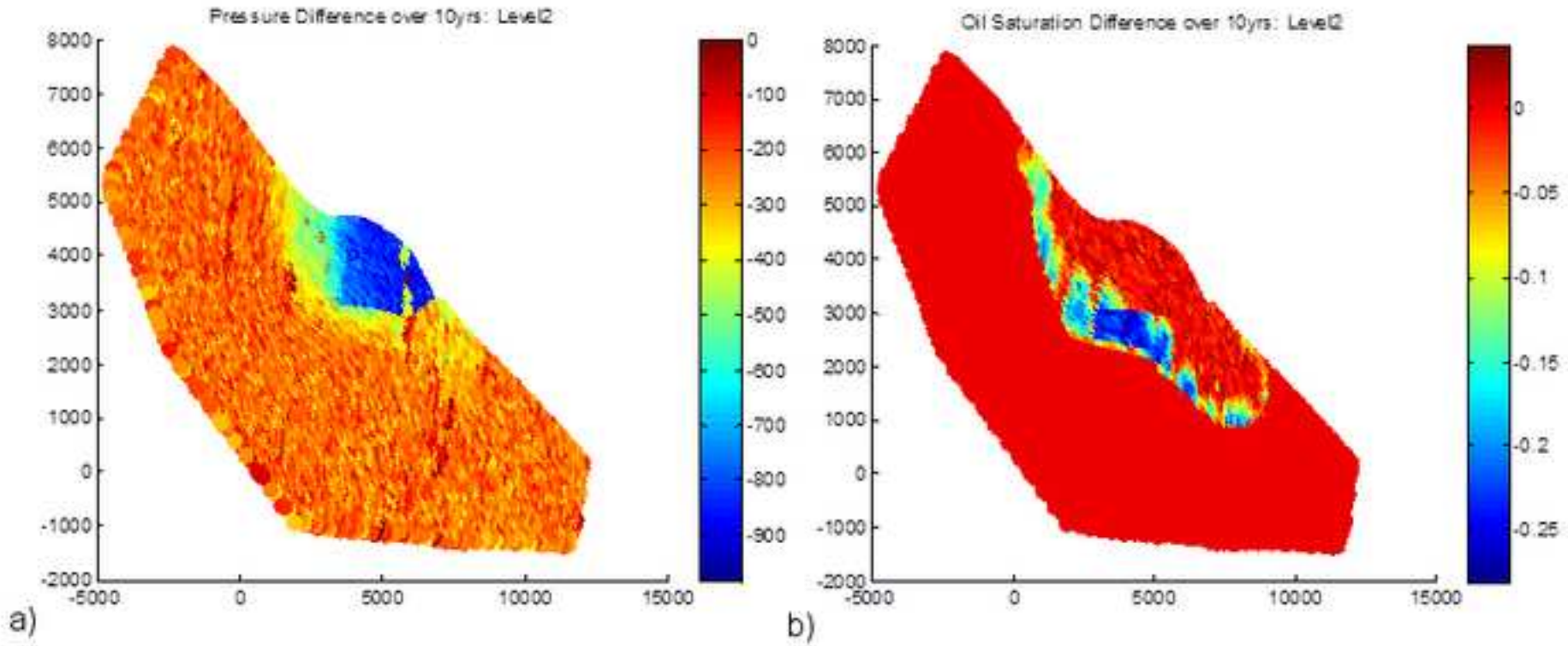

Figure 5

Click here to download high resolution image

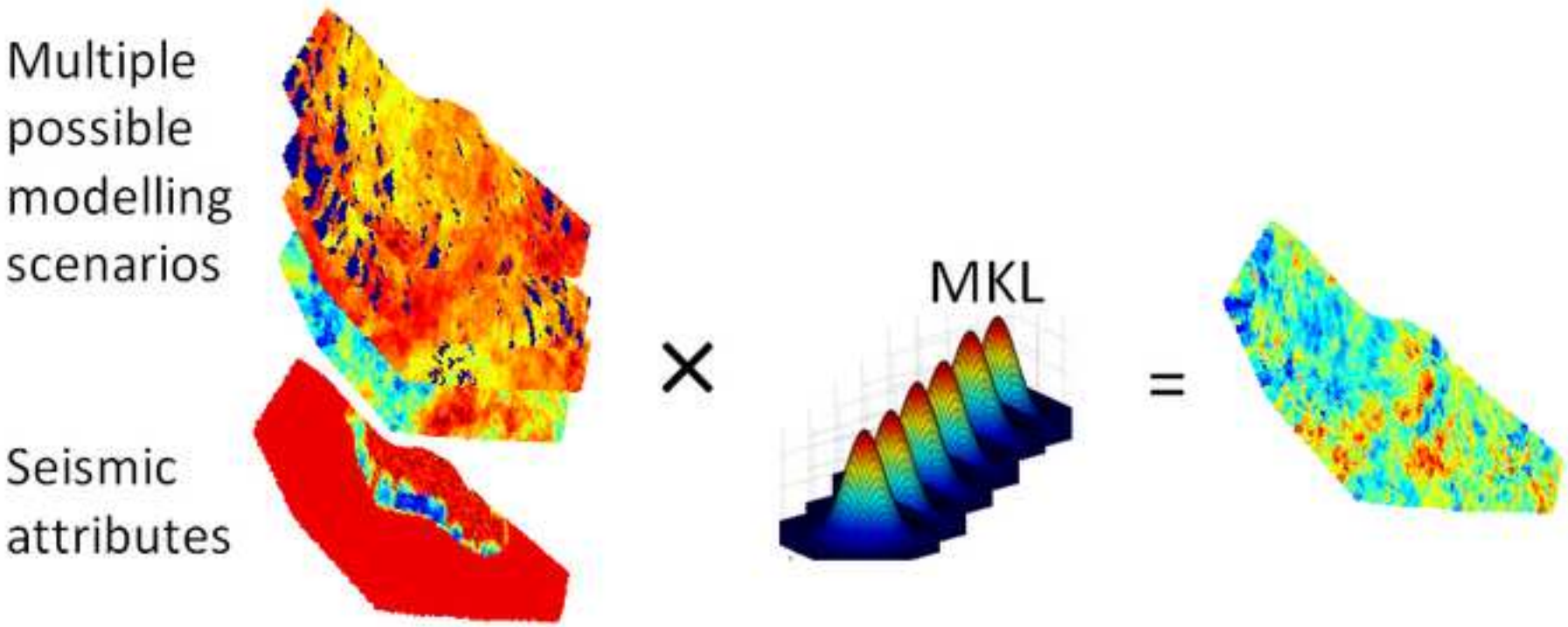

Figure 6

Click here to download high resolution image

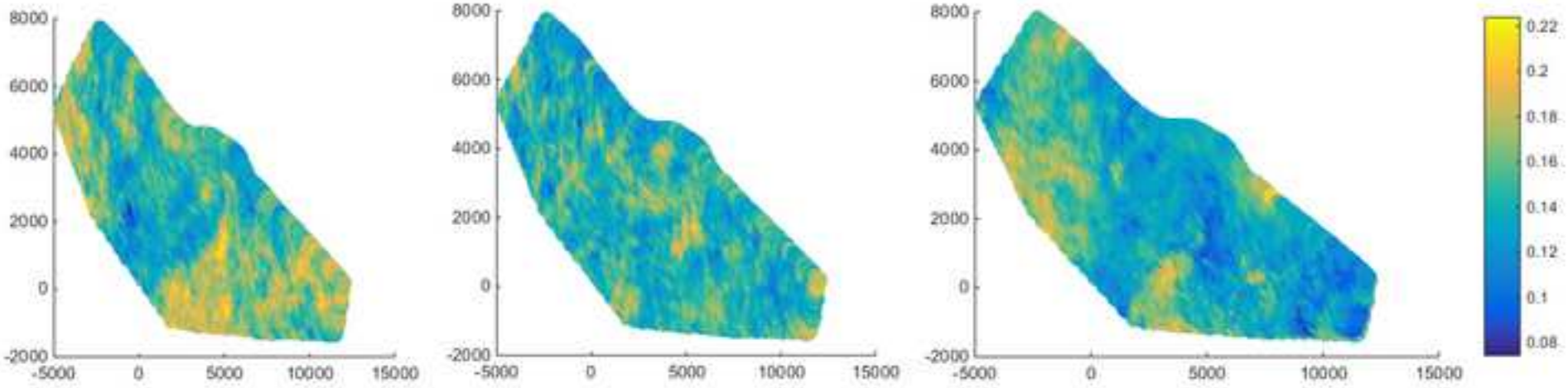

Figure 7

Click here to download high resolution image

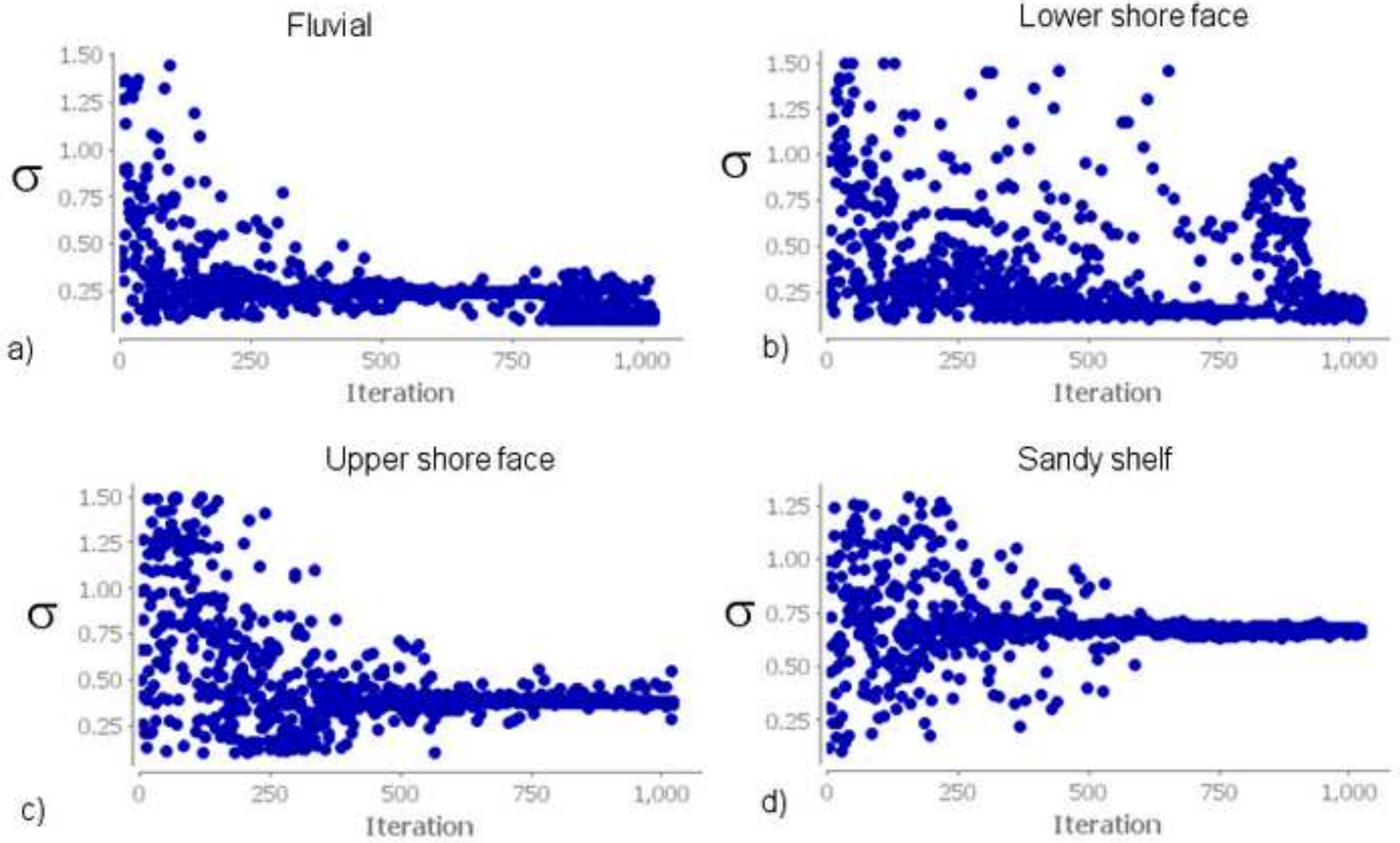

Figure 8a

Click here to download high resolution image

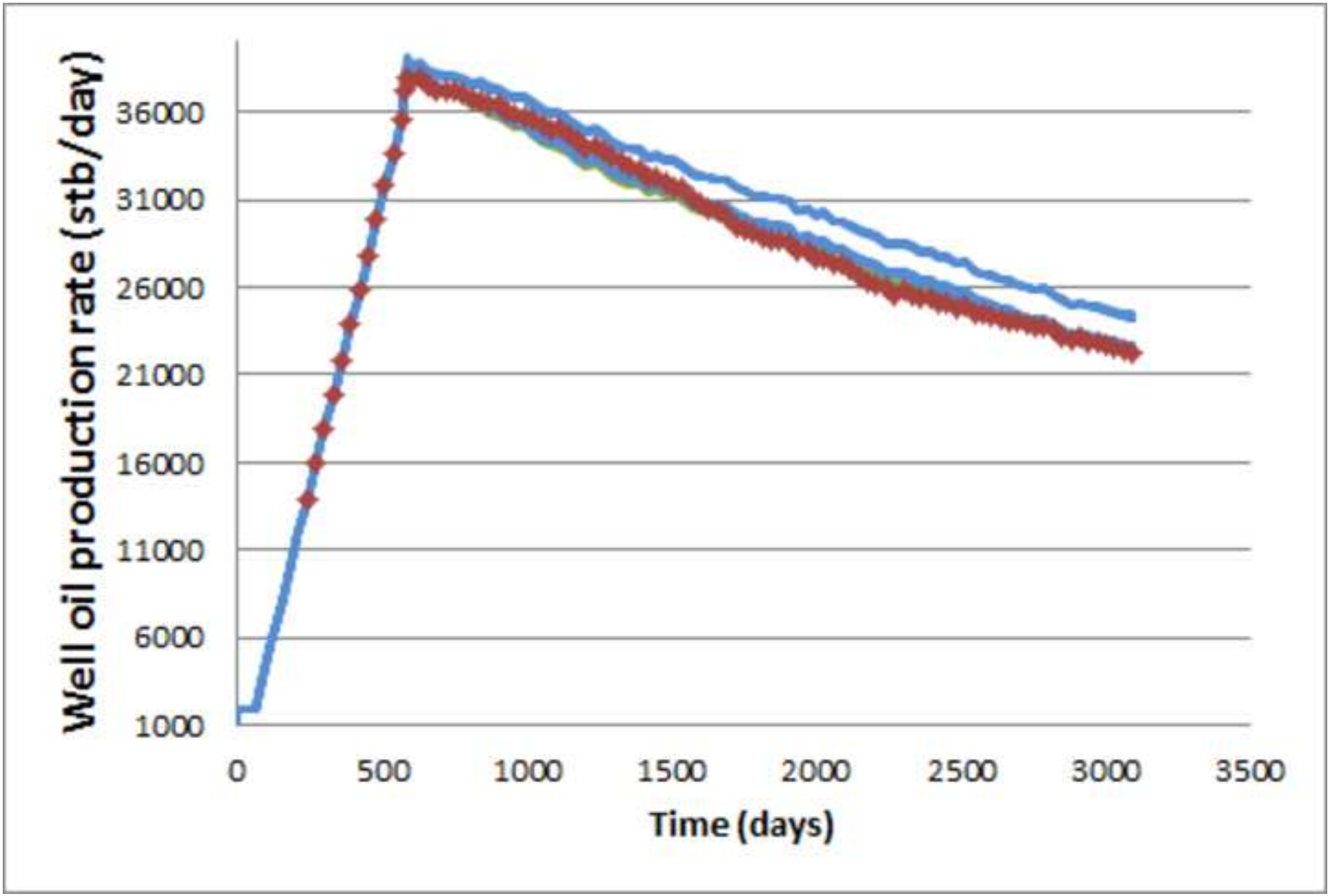

Figure 8b

Click here to download high resolution image

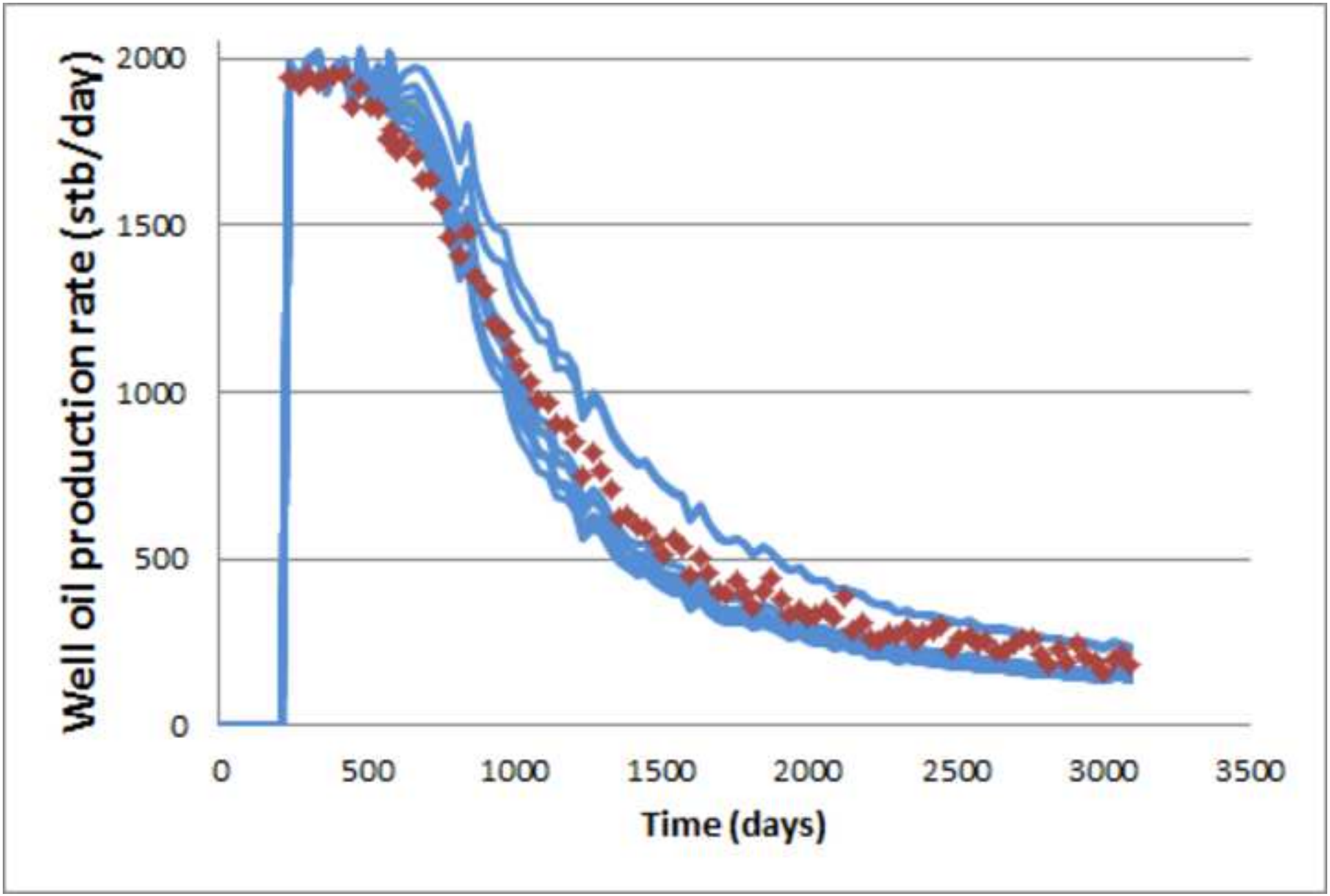

Figure 8c

Click here to download high resolution image

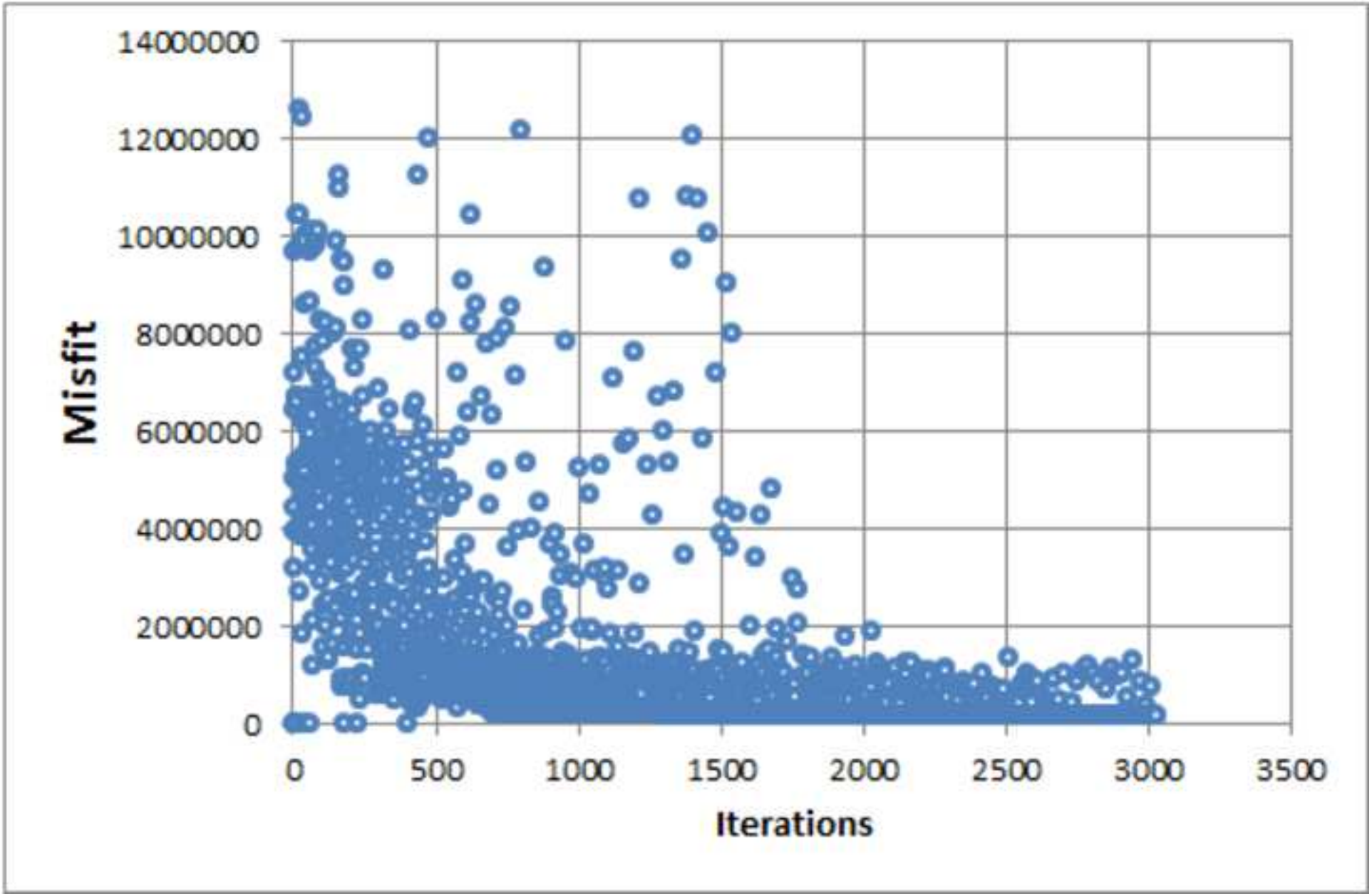

Figure 9a

Click here to download high resolution image

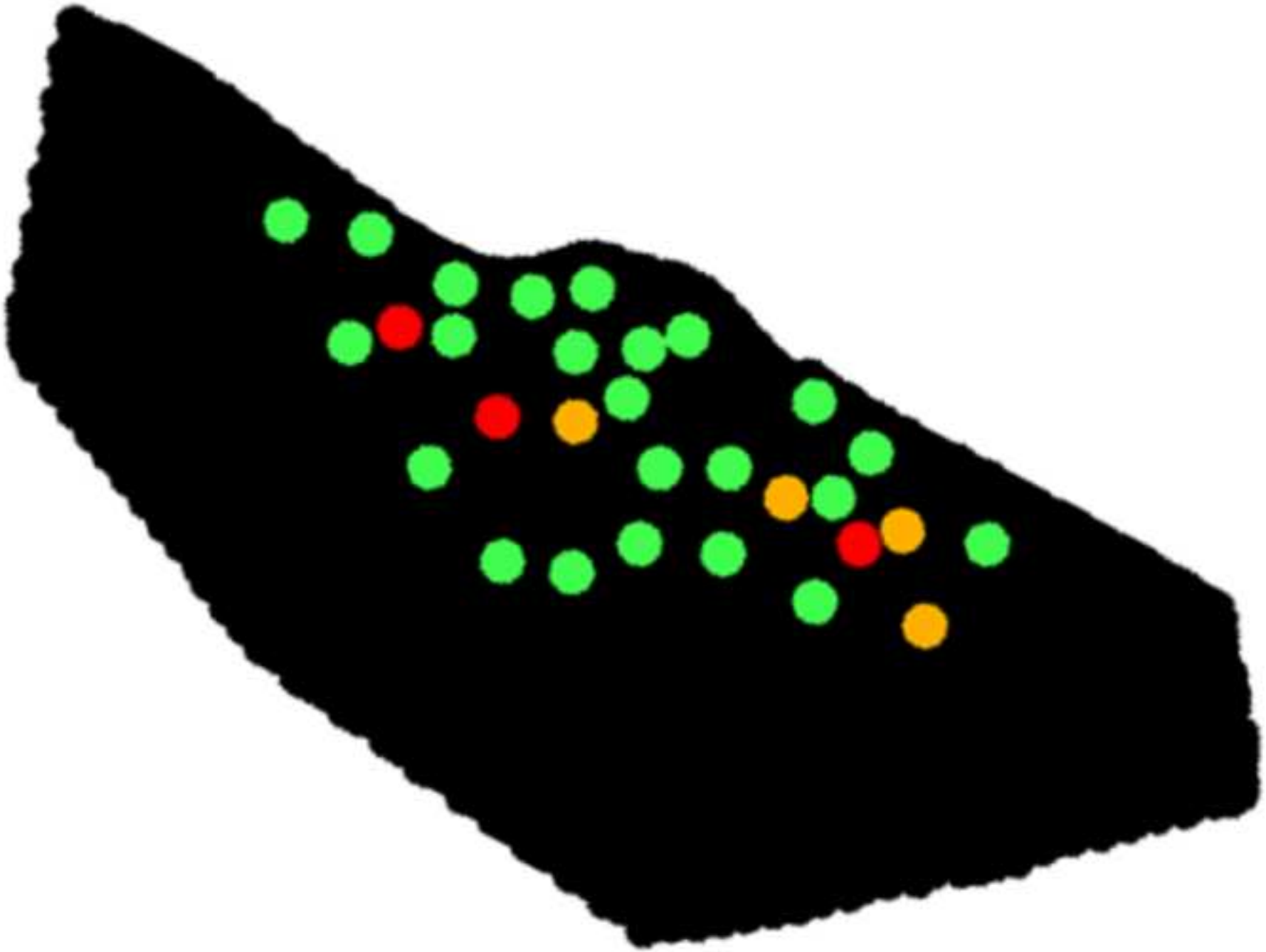

Figure 9b

Click here to download high resolution image

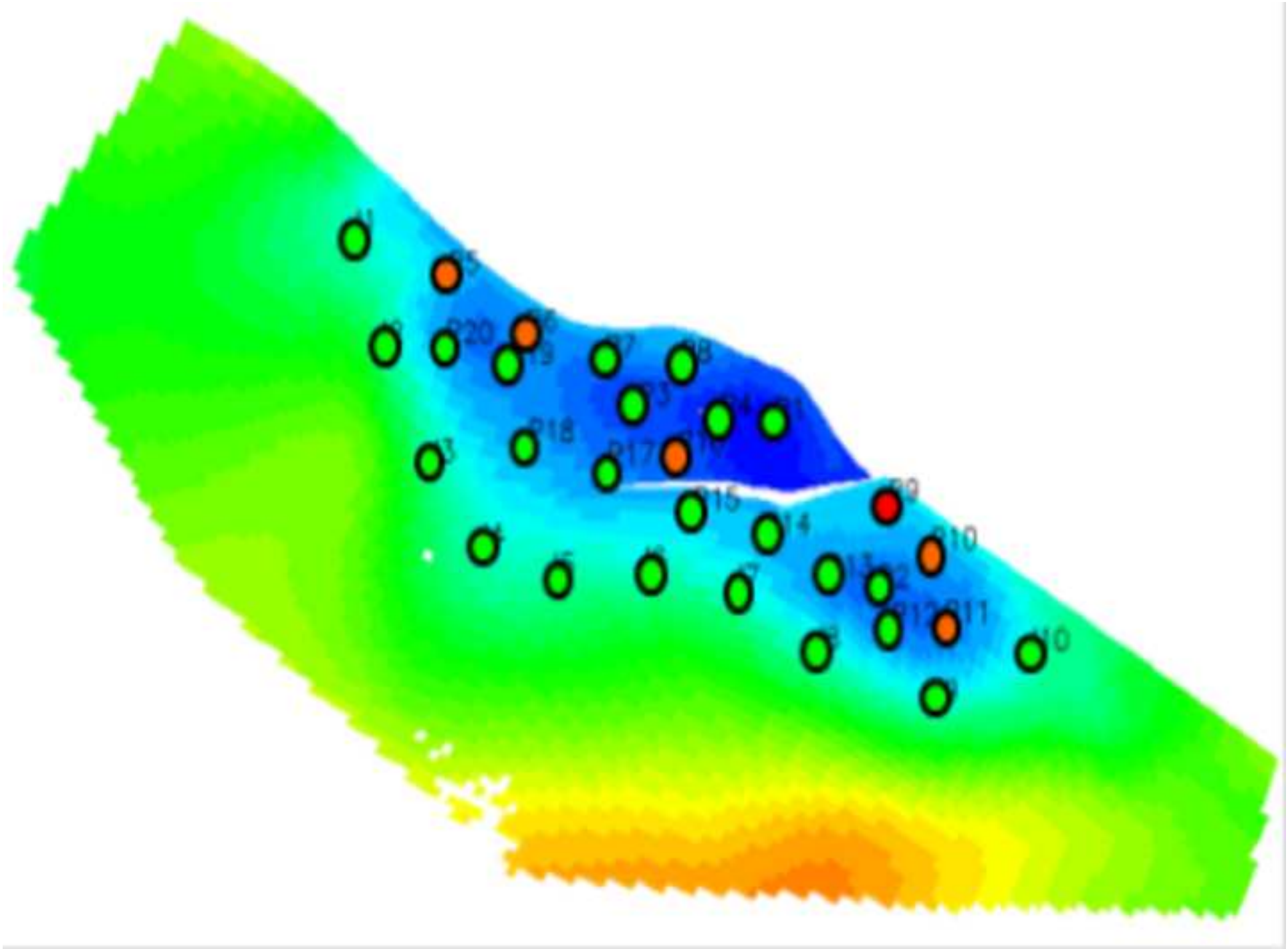

Figure 9c

Click here to download high resolution image

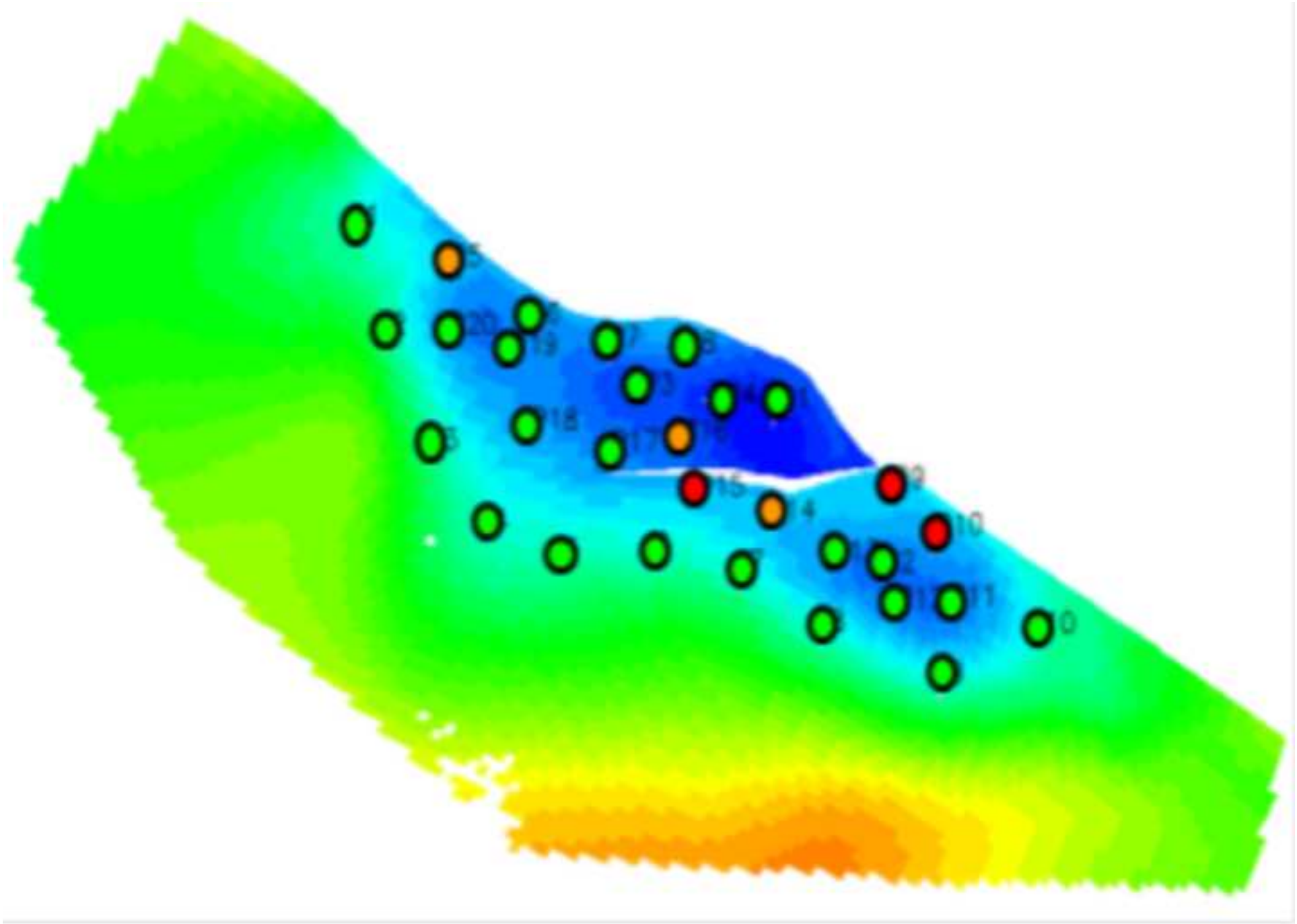

Figure 10

Click here to download high resolution image

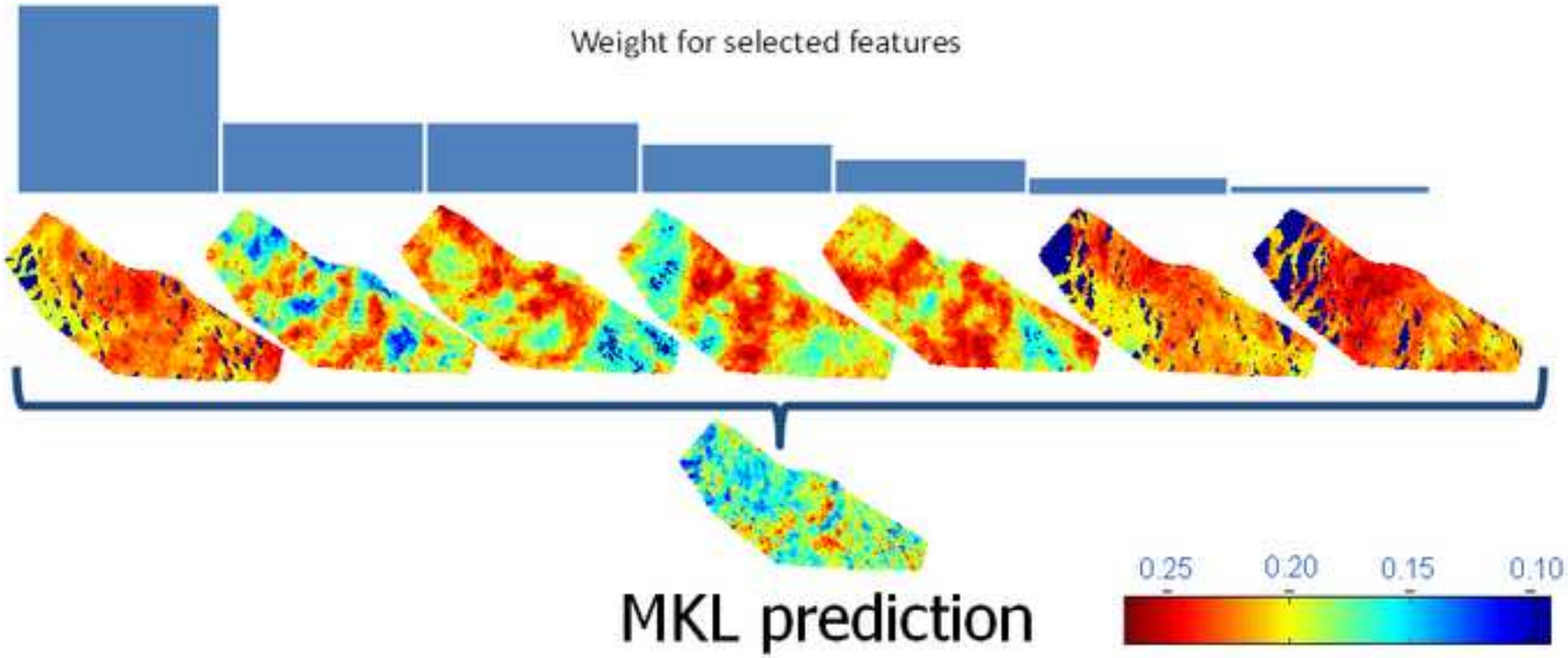

Figure 11

Click here to download high resolution image

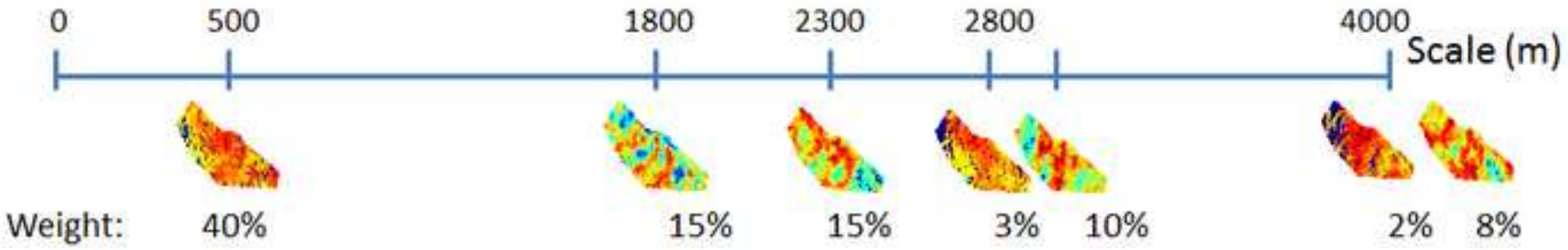

Figure 12

Click here to download high resolution image

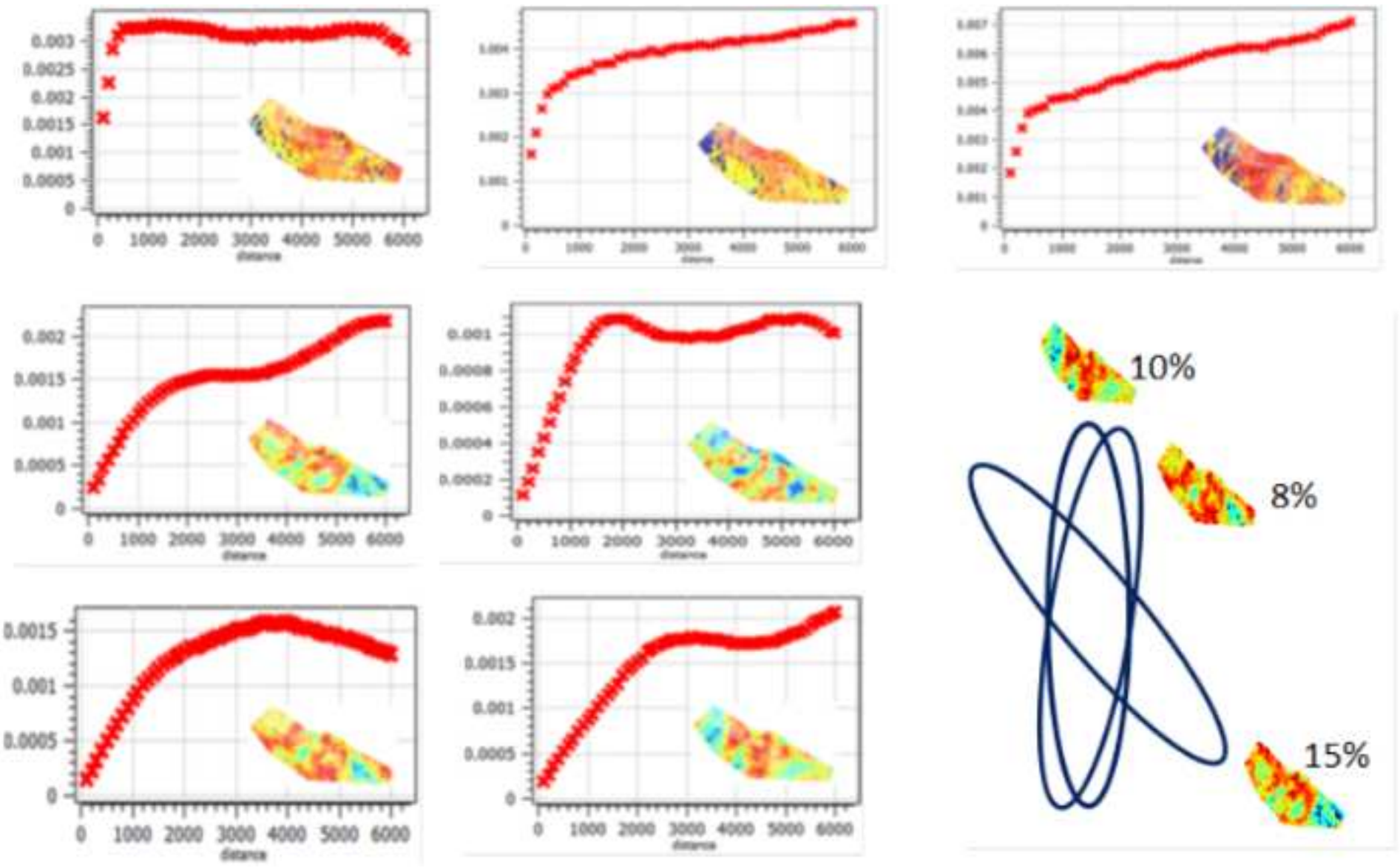

Figure 13

Click here to download high resolution image