Proceedings World Geothermal Congress 2020+1
Reykjavik, Iceland, April - October 2021

Machine Learning for Input Parameter Estimation in Geothermal Reservoir Modeling

[1]Anna Suzuki, [1]Megumi Konno, [2]Kimio Watanabe, [3]Kento Inoue, [4]Shinya Onodera, [4]Junichi Ishizaki, and [5]Toshiyuki Hashida

[1]Institute of Fluid Science, Tohoku University, 2-1-1 Katahira, Aoba-ku, Sendai, Miyagi, 980-8577 Japan

[2]Renergies Inc. 4-4-8 Kinshi, Sumida-ku, Tokyo, 130-0013 Japan

[3]Geothermal analysis, Inc., 7-15 Kauemon-cho, Tochigi, Tochigi, 328-0072 Japan

[4]Tohoku Electric Power Co., Inc., 1-7-1, Honcho, Aoba-ku, Sendai, Miyagi 980-8550 Japan

[5]6-6-11-707 Aza-Aoba, Aoba-ku, Sendai, Miyagi, 980-8579 Japan

anna.suzuki@tohoku.ac.jp

Keywords: natural state, reservoir modeling, simulation, TOUGH2, inverse analysis, automation, support vector machine

ABSTRACT

One of challenges in geothermal development is its uncertainty of estimation of complicated reservoir structures. For instance, permeability varies in several orders, and it is difficult to determine the distributions in a reservoir model. Although there are sophisticated inverse analysis methods (e.g., iTOUGH 2), people often determine the permeability distributions by trial and error according to their experiences. In the early stages of development, the estimation based on people's trial and errors are sometime quicker and more effective. If the process of permeability estimation is automated like the people's intuition, it would be useful to reduce uncertainty of modeling as well as to reduce simulation time and costs. In this study, we proposed a method to estimate permeability distributions by using measurement data based on machine learning. Several permeability distributions were given to numerical simulator, TOUGH 2, which generated the temperature and the pressure data as synthetic data. Combinations between the permeability and the temperature/pressure data were studied by support vector machine (SVM). The results shows the feasibility of estimating permeability distributions based on machine learning.

1. INTRODUCTION

Research on artificial intelligence (AI) is extremely active recently, and the AI-based research is applied to resource development and also geothermal development. The transition of the accumulated number of academic papers using “Artificial Intelligence” in geothermal fields is shown in Fig. 1. Figure 1 also shows the accumulated number of papers using “Machine Learning”, “Deep Learning”, “Artificial Neural Network”, “Geostatistics”, “Data-Driven”, “Data Assimilation”, and statistical and probabilistic approaches, such as “Bayesian Network / Theory”. Researches using “Artificial Intelligence” surged between 2005 and 2010, which catches up those using geostatistics (Google scholar, https://scholar.google.com/). As shown in Fig. 1(b), “Deep learning” has become popular words with geothermal.

(a) **(b)**

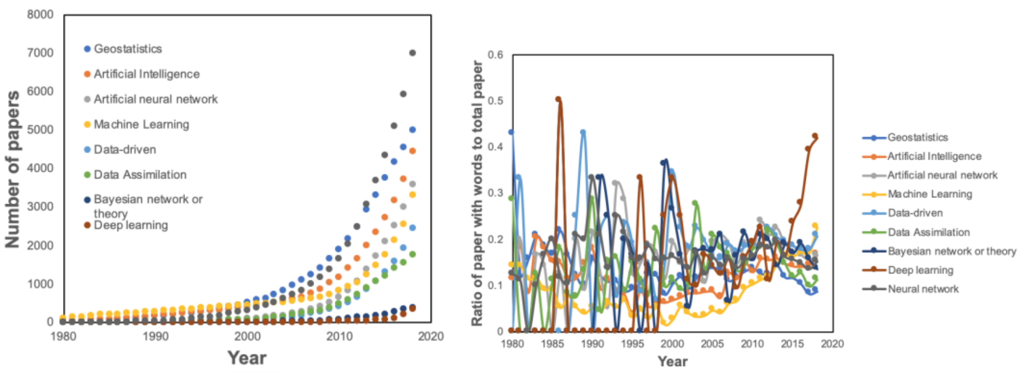

Figure 1: Accumulated number of journal papers using the words in geothermal fields (Google scholar, https://scholar.google.com/): (a) cumulated number of papers and (b) ratio of papers with the words to total papers.

In this study, we will use the AI technology to geothermal reservoir modeling. Schematic of conventional and new approach of geothermal modeling for natural state is shown in Fig. 2. For numerical modeling of geothermal reservoirs, physical properties of the rock are distributed to each grid on simulation area based on the conceptual models including the geological information. Governing equations of fluid flow and heat transfer are solved in each grid to determine the behavior of the entire reservoir. The simulation results (i.e., temperature and pressure) are compared to the observed data and the physical properties are adjusted to fit to the observed data. If there are fractures, permeability changes by several orders. So, the permeability is an important parameter that greatly affects the results of the numerical simulation. Although there are some inverse approaches, such as iTOUGH2 (Finsterle, 2007), the permeability distributions are estimated by trial and errors based on experiences of analysists. Because the number of combination of parameters are extremely large, it takes quite long time and is easy to converge to the local optimized values.

To automate the estimation of permeability distributions using well data, we proposed an AI-based approach of geothermal reservoir modeling. We will develop a model to estimate permeability distribution using well data directly. Yet, in this paper, we considered that the 3D temperature and pressure distribution can be obtained by Kriging (e.g., Koike et al., 2001). The 3D temperature and pressure distributions were used for the estimation.

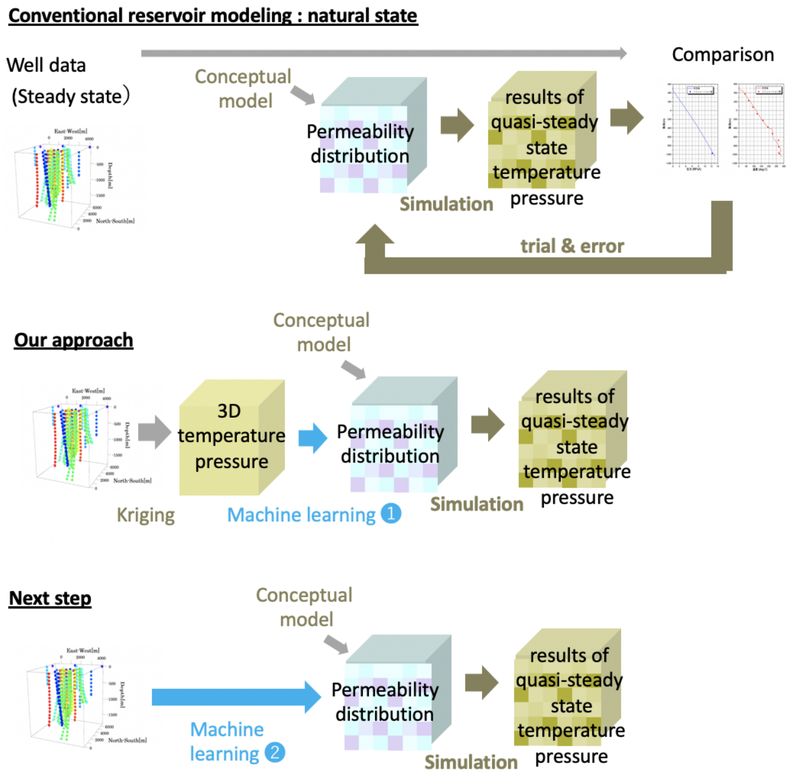

Figure 2: Schematic of geothermal reservoir modeling and our approach.

2. METHOD

In this study, we tried to learn the relationship between the permeability distribution and the temperature and pressure distribution based on machine-learning algorithms and to estimate the permeability from the temperature and pressure distribution (i.e., inverse analysis).

2.1 Forward analysis

Numerical simulator TOUGH2 (Pruess et al., 2015) was used to simulate three-dimensional fluid and heat flow. The mass conservation equation and energy balance equation were used in TOUGH2. The conceptual diagram of the model is shown in Fig. 3, and the simulation conditions are shown in Table 2. As shown in Fig. 3, the simulation area was 34 grids x 26 grids x 17 layers. The top was open boundary with temperature of 30°C and pressure of 2.5 MPa. The bottom boundary was no flow condition. A hot water source was installed on the bottom boundary as shown in Fig. 3.

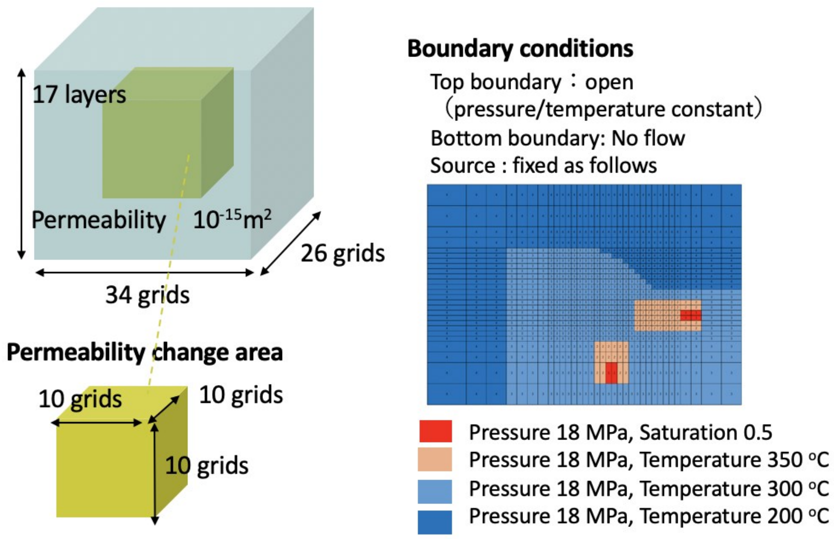

Figure 3: conceptual simulation model in TOUGH2.

Table 1: Simulation condition.

Parameters	**Values**
Density [kg/m^3]	2250.0
Porosity: surrounding rocks	0.01
flow paths	0.1
reservoir	0.1
Heat conduction []	2.5
Specific heat []	1000
Porosity: surrounding rocks [m^2]	10^{-20}
flow paths [m^2]	10^{-14}
reservoir [m^2]	10^{-16}-10^{-11}
Source flow rate [kg/s]	0.1
Source enthalpy [J]	1.085 x 10^6

The initial permeability was 10^{-15} m^2. The permeability in the yellow area (10 x 10 x 10 grids) was changed using the discrete cosine transformation (Ahmed et al., 1974) given by

$$X_{k_1,k_2,k_3} = \sum_{n_1=0}^{N_1-1}\sum_{n_2=0}^{N_2-1}\sum_{n_3=0}^{N_3-1} x_{n_1,n_2,n_3} \cos\left[\frac{\pi}{N_1}\left(n_1+\frac{1}{2}\right)k_1\right]\cos\left[\frac{\pi}{N_2}\left(n_2+\frac{1}{2}\right)k_2\right]\cos\left[\frac{\pi}{N_3}\left(n_3+\frac{1}{2}\right)k_3\right], \quad \forall k_i = 0,1,2,\ldots,N_i-1. \tag{1}$$

The discrete cosine transform is a basic image generation method used for image analysis. By using the function, we could generate many permeability distributions automatically. One of example for the result of generated discrete cosine transformation and simulation results of TOUGH2 are shown in Fig. 4.

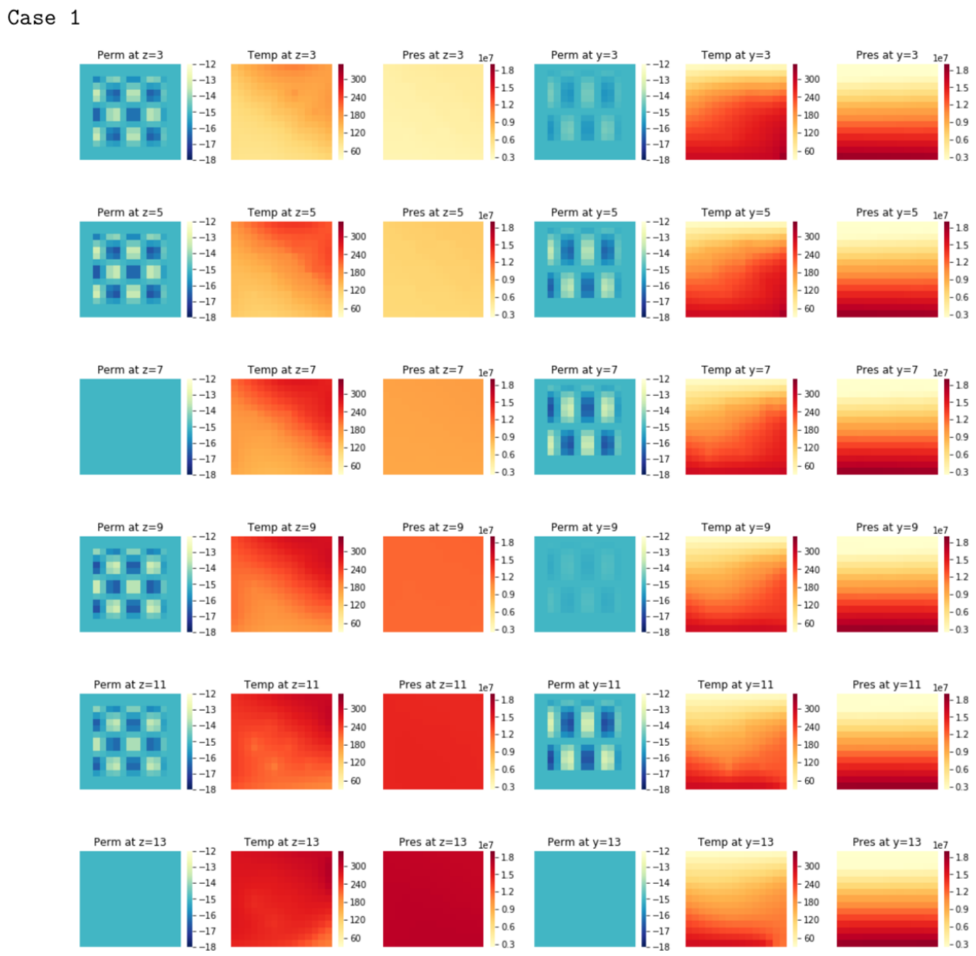

Figure 4: Permeability distributions generated by discrete cosine transformation and numerical results from TOUGH2.

2.2 Inverse analysis

The permeability is estimated based on temperature and pressure distribution calculated by TOUGH2. First, the data set was prepared. The predicted value is called the "target". We set the value of the permeability at each grid to the target, which is represented as a one-dimensional array $\boldsymbol{y}$. The length of $\boldsymbol{y}$ is the number of grids of each simulation M times the number of simulations N (M x N). The input parameters are called the "feature", which is represented as $\boldsymbol{X}$ with the size of (M x N, the number of feature quantities). We used the temperature gradient and the pressure gradient at the target grid as the feature quantities as shown in Fig. 5. The gradient was calculated with the difference between the values at the target grid and those of one or three neighboring points. The predicted value $\boldsymbol{y}^*$ is calculated by learning the combination of $\boldsymbol{y}$ and $\boldsymbol{X}$.

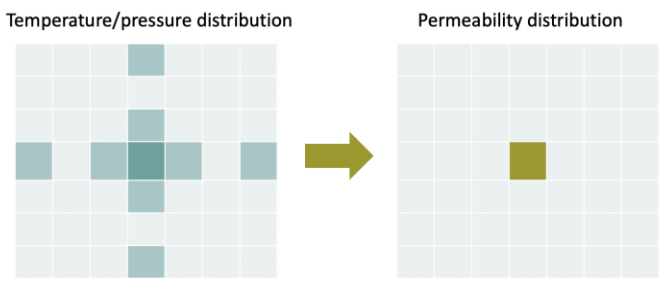

Figure 5: Feature data set (temperature and pressure gradient) and target data set (permeability).

In this study, we used scikit-learn in Python to develop the machine learning algorithm (Pedregosa et al., 2011). The learning was based on regression analysis. Regression analysis is a statistical technique that is performed to analyze trends in two sets of data, and one of two variables that may be correlated or causally related is used for the others. This is classified using a support vector machine (Smola and Schölkopf, 2004). The support vector machine is a classification method that searches for boundaries separating data with as wide margin as possible and detects the best boundary if it can not be clearly separated.

First, normalize the data set with sklearn.preprocessing.MinMaxScaler (). In this way, it is possible to prevent information from being lost even if the scale of feature quantities are different in digit. The data is divided into training data (X_{train}, y_{train}) and test data (X_{test}, y_{test}) by sklearn.cross_validation.train_test_split (). In this study, 30% of data was used as test data, and 70 % of data was used as training data. Subsequently, the optimal value of the hyper parameter was determined using sklearn.model_selection.GridSearchCV (). With support vector machines, the radial basis functions are the kernel, so the hyperparameters were *gamma* and *C* in this case. The mean squared error was used for the evaluation function. In addition, cross validation was performed five times using sklearn.cross_validation.cross_val_score () to evaluate the combination of training data and test data. After that, training was performed by fit () using (X_{train}, y_{train}), and y_{test}* was estimated by predict using X_{test}.

3. RESULTS

The estimation results using training data and test data are shown in Figs. 6 and 7, respectively. The left figures are the expected permeability distribution of XY plane, which were used as the input data in TOUGH2. The right figures are predicted results of XY plane, which were obtained by machine learning using temperature and pressure data from TOUGH2. As shown in Fig. 6, the estimation could characterize the trend of distributions. Fig. 7 shows the estimation results using test data. Although the estimation was not good quality compared to using training data, the estimation was reasonable. Fig. 8 shows the accuracy of estimation. The x axis in Fig. 8(a) is the expected results and y axis is the predicted values. The trend is reasonably good. Fig. 8(b) shows the training curve. When increase in the ratio of training data, the estimation was improved. However, the accuracy is not improved when we used over 10000 training data set. Thus, the data set was good enough number, and we need to consider the different method to improve the accuracy.

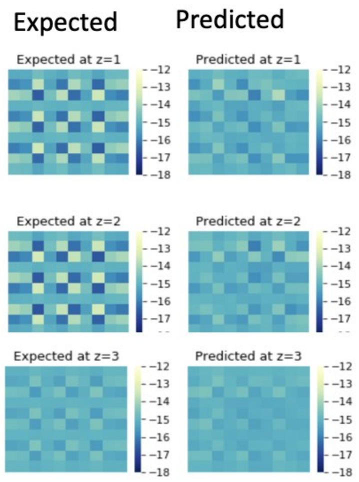

Figure 6: Estimation results for training data. Left figures are original permeability distribution, and right figures are predicted results by machine learning.

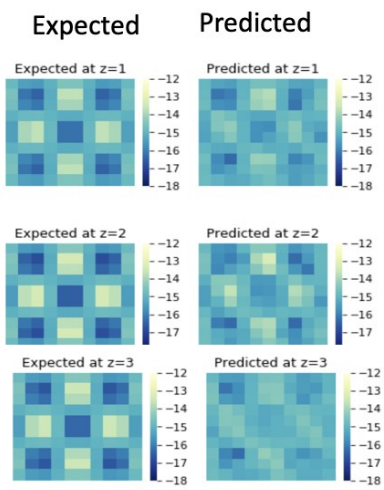

Figure 7: Estimation results for test data. Left figures are original permeability distribution, and right figures are predicted results by machine learning.

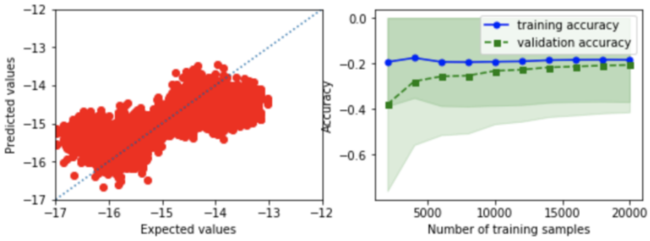

Figure 8: Accuracy of estimation for testing data. (a) expected values vs predicted values and (b) training learning curve.

We used more realistic permeability distribution to validate the machine learning method. Figure 9 shows the permeability distribution of XZ plane at the left-hand side. The permeability distribution was generated by assuming a natural geothermal field. The top area has shallow reservoir with high permeability and the bottom of the shallow reservoir has a cap rock. By using the permeability distribution, we simulated natural state and obtained temperature and pressure distribution from TOUGH2. As shown in Fig. 9, the prediction is reasonable. Fig. 10 shows the accuracy of estimation. The predicted values are close to the expected values. Even there are the spatial heterogeneity, the machine learning algorithm could estimate the permeability distributions.

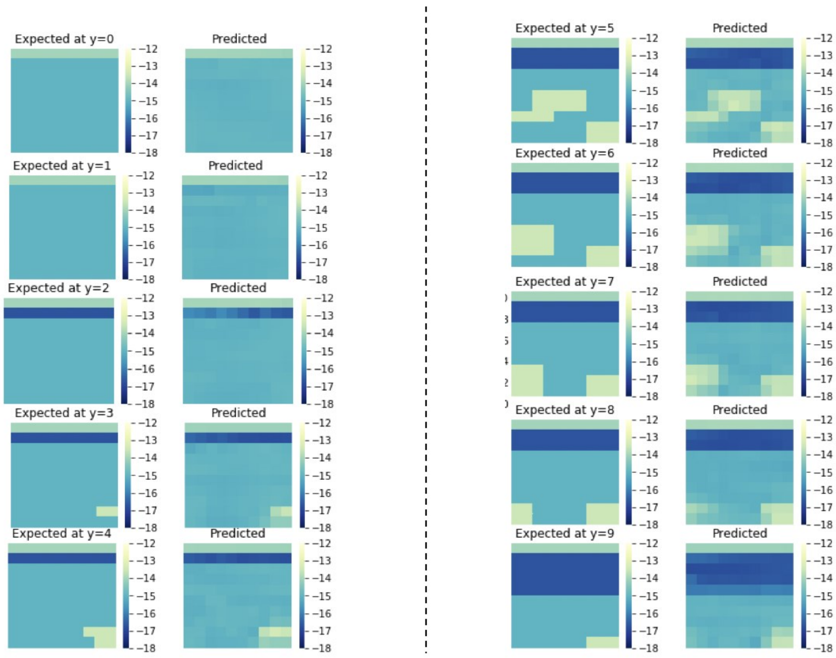

Figure 9: Estimation results of synthetic data assuming a natural geothermal reservoir. Left figures are original permeability distribution, and right figures are predicted results by machine learning.

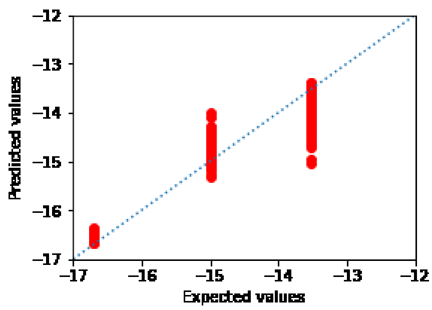

Figure 10: Accuracy of estimation for synthetic data assuming a natural geothermal reservoir.

CONCLUSIONS

We have proposed a method for estimating the permeability from the measured data automatically by applying machine learning to geothermal reservoir modeling. The learning data was produced by using TOUGH2. The permeability distribution was created based on the discrete cosine transformation, and the temperature and the pressure were simulated in TOUGH2. The characteristics of the combination between the permeability at each grid and the gradients of the temperature and the pressure was learned by the support vector machine (SVM). Our results show that the permeability distribution assumed an actual reservoir was predicted accurately. In the next step, it is desirable to update the method that can be estimated not from 3D data but from well data directly and to consider history matching (unsteady state conditions). The proposed method can estimate the permeability distribution without consuming time and without depending on analyst's experience, which will be a great technology to advance the geothermal reservoir modeling.

REFERENCES

We would like to thank Prof. Fukui, the Institute of Scientific and Industrial Research, Osaka University for his kind discussion and advices. Requests for numerical data and algorithms should be addressed to Anna Suzuki (e-mail: anna.suzuki@tohoku.ac.jp).

REFERENCES

Ahmed, N., Natarajan, T., and Rao, K. R.: Discrete cosine transform, *IEEE transactions on Computers*, **100**(1), (1974) 90-93.

Finsterle, S., iTOUGH2 User's Guide, Report LBNL-40040, Lawrence Berkeley National Laboratory, Berkeley, California (2007).

Koike, K., Matsuda, S., and Gu, B: Evaluation of Interpolation Accuracy of Neural Kriging with Application to Temperature-Distribution Analysis, *Mathematical Geology,* **33**(4), (2001), 421-448.

Pruess, K., C. Oldenburg, and G. Moridis, TOUGH2 User's Guide, Version 2, Report LBNL-57952, Lawrence Berkeley National Laboratory, Berkeley, California (2015).

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al.: Scikit-learn: Machine learning in Python. *Journal of machine learning research*, **12**(Oct), (2011), 2825-2830.

Smola, A. J., and Schölkopf, B.: A tutorial on support vector regression. *Statistics and computing*, **14**(3),(2004), 199-222.