

Points Property Prediction

Points property prediction using Lineal regression, ACE regression, Neural Network regression or Random Forest regression

User Manual

IPLAB

Copyright © 2016 IPLAB. All rights reserved.

This work contains the confidential and proprietary trade secrets of IPLAB and may not be copied or stored in an information retrieval system, transferred, used, distributed, translated or retransmitted in any form or by any means, electronic or mechanical, in whole or in part, without the express written permission of the copyright owner.

Table of Contents

1. Introduction	3
2. Start	4
3. Input parameters	5
1. Calculation	8

1. Introduction

The **Points Property Prediction** IP_Seismic plug-in (version: 2017.1.0.0, release date: July 2017) can be used to predict a dependent variable (response) using one or more independent variables (predictors or covariates).

To do the prediction there are two main calculation studies:

1. Training stage – to get prediction operator (coefficients of the operator) according set of training pairs (one response and one or more predictors). Prediction operator can be Linear Regression, ACE regression (Alternating Conditional Expectation, LEO BREIMAN and JEROME H. FRIEDMAN, 1985), Neural Network regression or Random Forest regression.
2. Calculation stage – to calculate values of response based on prediction operator for all predictors.

2. Start

start: Surface Attributes->

Surface Property Prediction

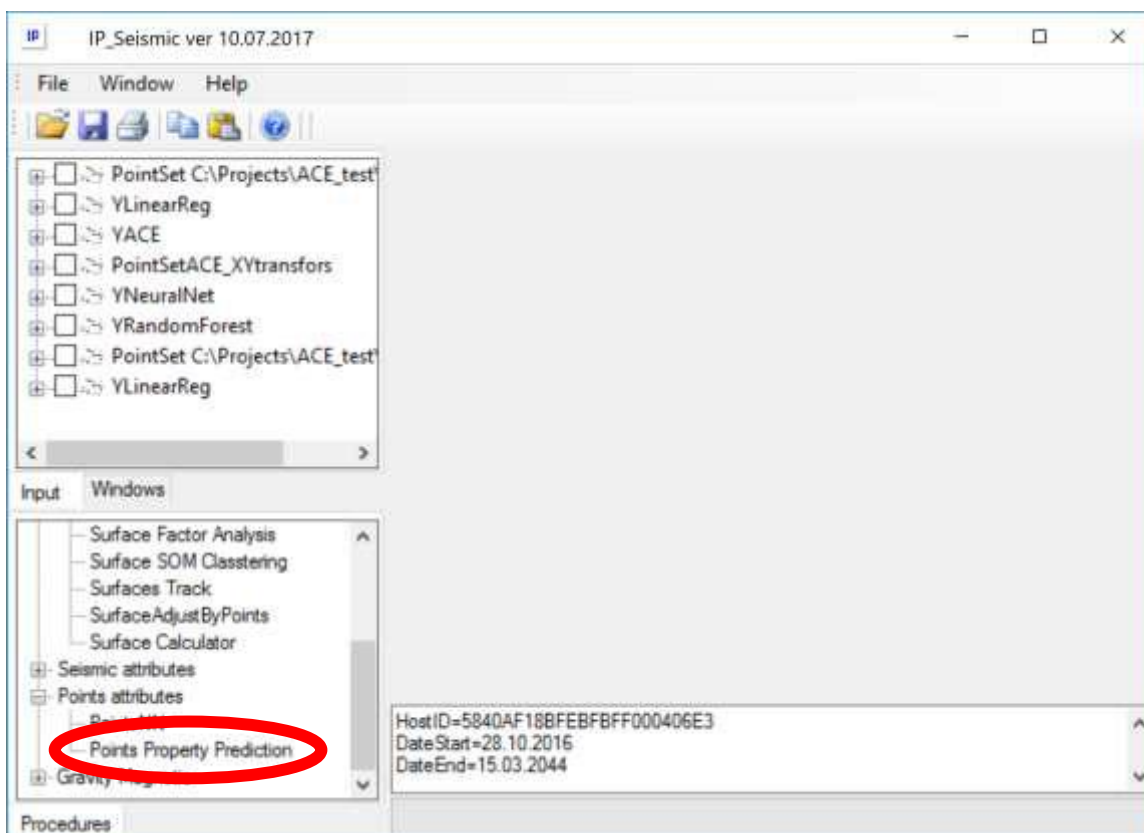


Figure 1: Project tree and programs tree to start **Surface Property Prediction**

3. Input parameters

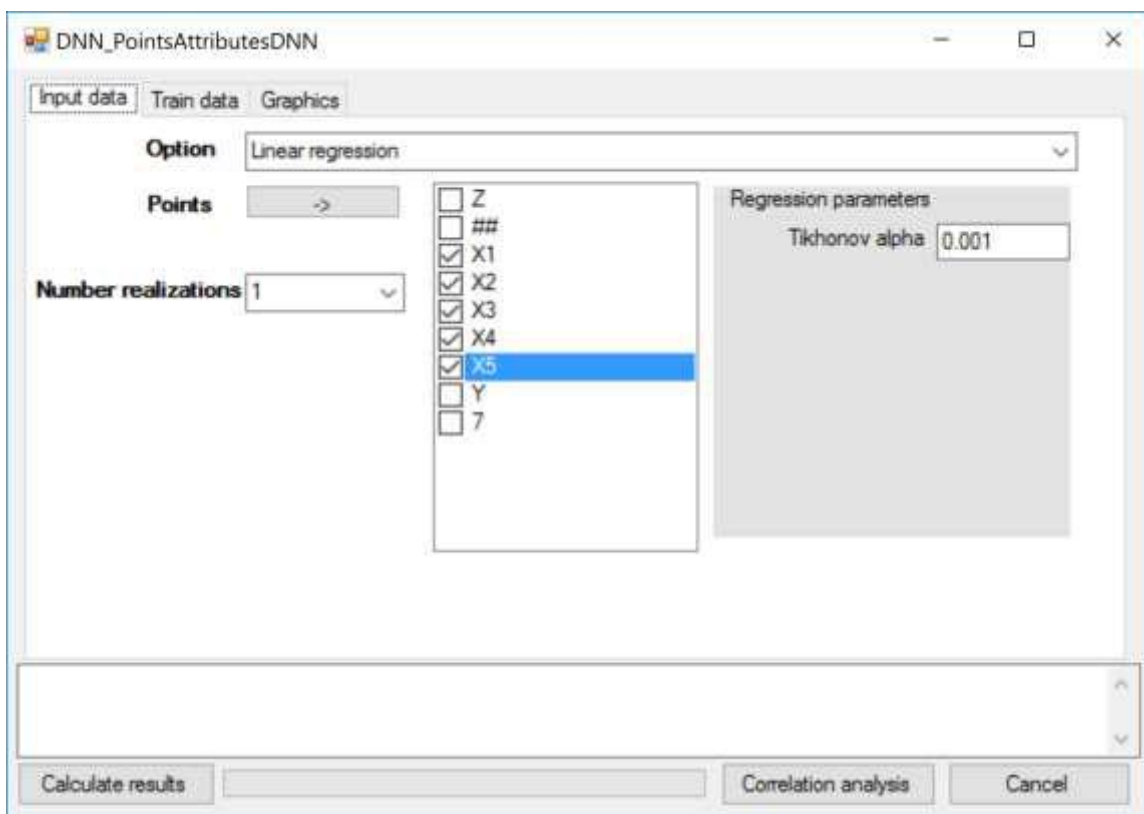


Figure 2: Input surface attributes data tab dialog view **Points Property Prediction**

Parameters have to be defined before calculation:

Points: allow select points with set of attributes from the project tree. All marked attributes will be used for calculations like Predictors.

Option:

Linear Regression: prediction based on Linear Regression.

ACE Regression: prediction based on ACE Regression

Neural Network: prediction based on Neural Network Regression

Random Forest: prediction based on Random Forest Regression.

Parameters for different options:

Tikhonov alpha: >0 and <1 allow avoid overlearning effect or instability for prediction. If $\alpha=0$ then we can get very good approximation of the training set but the predictability can be very low and results can be very different for every

realization. If $\alpha > 0$ then training quality (correlation coefficient) will be less if use $\alpha=0$, but predictability will be much higher (quality control with “blend well” test).

Iteration: allow define maximum iteration during training stage to teach the neural network.

Cross validation (%): (≥ 0 and ≤ 99) allow define percent of training pairs what will not be used for training and will be used only for quality control.

Deep learning option: Not used default, **Kohonen SOM, Restricted Boltzmann machine . Autoencoder.** Option allow to generalized first layers.

Hidden layer 1 (2, 3, 4, 5) nodes: allow to define number nodes in 1st (2nd, 3rd, 4th, 5th) hidden layer for neural network definition.

To select production points attributes need to use second tab:

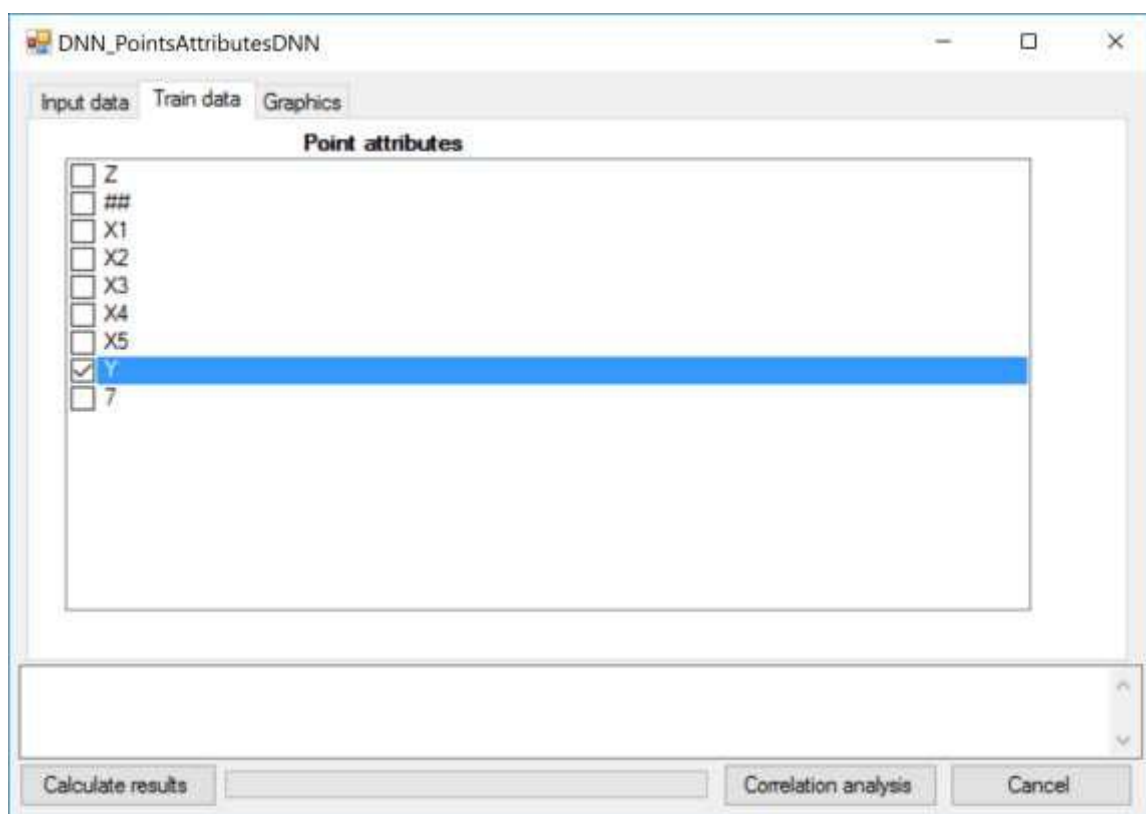


Figure 3: Response attribute from points attributes tab dialog

To do correlation analysis need to push “Correlation Analysis” bottom. After calculation you can see correlation table (see Figure 4) with cress correlation values

and with Principal components (shift table to right). Table is sortable for every column (click to column name). Double click – sort from big to small.

Correlation Analysis												
Please original sorting		Number points used for calculation										
	#	Mean	Std	Y	X1	X2	X3	X4	X5	Factor1 weight= 25.50%, sum= 49.54%	Factor2 weight= 15.64%, sum= 45.54%	Factor3 weight= 15.62%, sum= 85.36%
*		0.68786046	0.00067511	1.00000000	0.21548949	-0.15341622	-0.00020967	0.29843557	0.53134481	0.5142	0.2912	-0.0905
	X1	-0.09223000	0.55668388	0.21548949	1.00000000	-0.12776739	0.09957899	0.05185208	-0.08620495	0.2671	-0.3420	-0.6434
	X2	-0.07114978	0.55474466	0.15341622	-0.12776739	1.00000000	-0.16253894	-0.17418917	-0.15996715	-0.3738	0.3727	-0.1874
	X3	-0.00636182	0.56159257	0.00020967	0.09957899	-0.16253894	1.00000000	0.09374237	0.01386396	0.1598	-0.8205	0.1816
	X4	-0.02806181	0.56518073	0.29843557	0.05185208	-0.17418917	0.09374237	1.00000000	-0.02686462	0.3868	-0.2318	0.4659
	X5	0.00026124	0.51296423	0.53134481	-0.08620495	-0.15996715	-0.01386396	0.02686462	1.00000000	0.4583	0.4700	0.0383
*												

Figure 4: Correlation table results according Correlation Analysis bottom

4. Calculation

To start the prediction need to push "Calculate result" button (during training stage error and correlation curves will be calculated in Graphics window (see Figure 5))

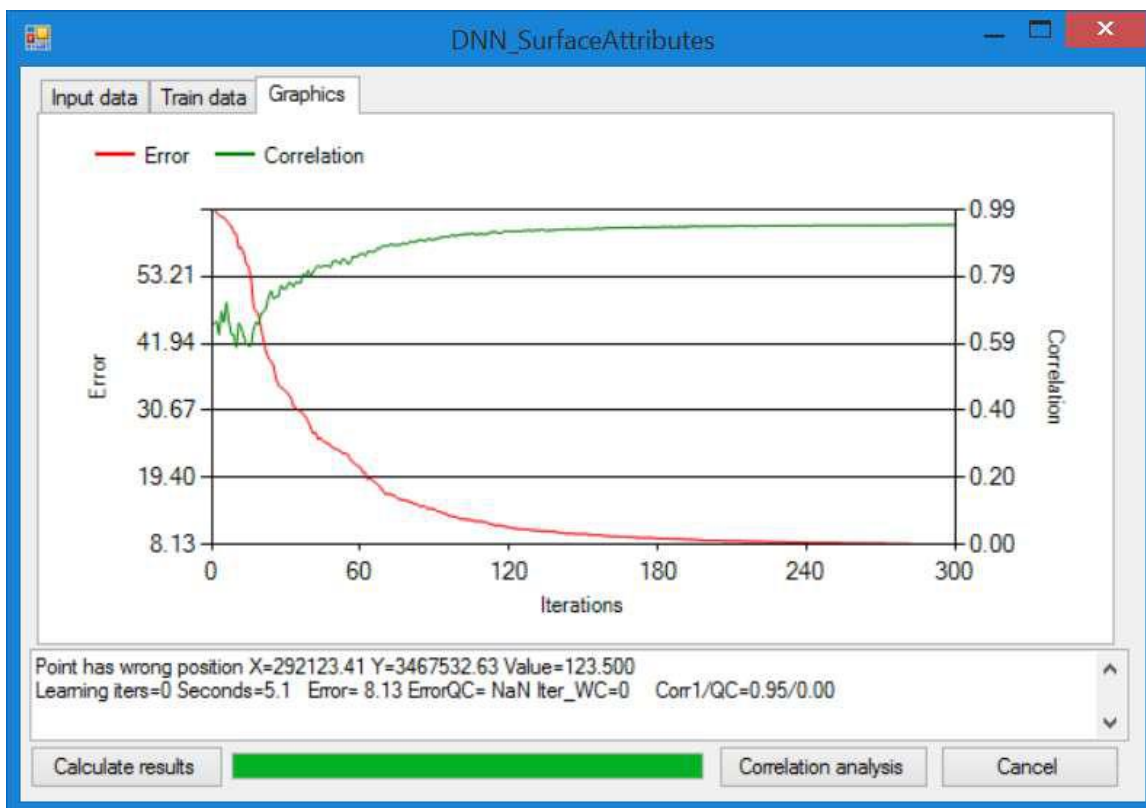


Figure 5: Output graphics window (only for Neural Network option) with error and correlation values during learning iterations

5. Results

As results of the calculations will be only new points set with measured and predicted response variable.

For ACE regression option will create new point sets with Transforms functions for every used variables (response and predictor)

6. Test

To test the procedures, we can recommend doing the following test workflow (D. Wang and M. Murphy, 2004, Estimating Optimal Transformations for Multiple Regression Using the ACE Algorithm):

1. Create Excel worksheet with several columns

X1	X2	X3	X4	X5	Y
-.342749805	-.560210799	0.023656124	0.171211726	-.618720465	0.478112647
-.145980504	-.362085949	0.786882801	-.262804618	-.078953371	0.64075565
0.035345863	0.507438205	-0.02095473	0.988718677	0.950364055	0.396192053

.....

Where X1, X2, X3, X4, X5 are random (-1,1) variables and

$$Y = \log[4 + \sin(4X1) + |X2| + X2^3 + X3^4 + X5 + .1]$$

The Transform function for it will be following:

$$\theta^*(Y) = \exp(Y)$$

$$\varphi^*_1(X1) = \sin(4X1)$$

$$\varphi^*_2(X2) = |X2|$$

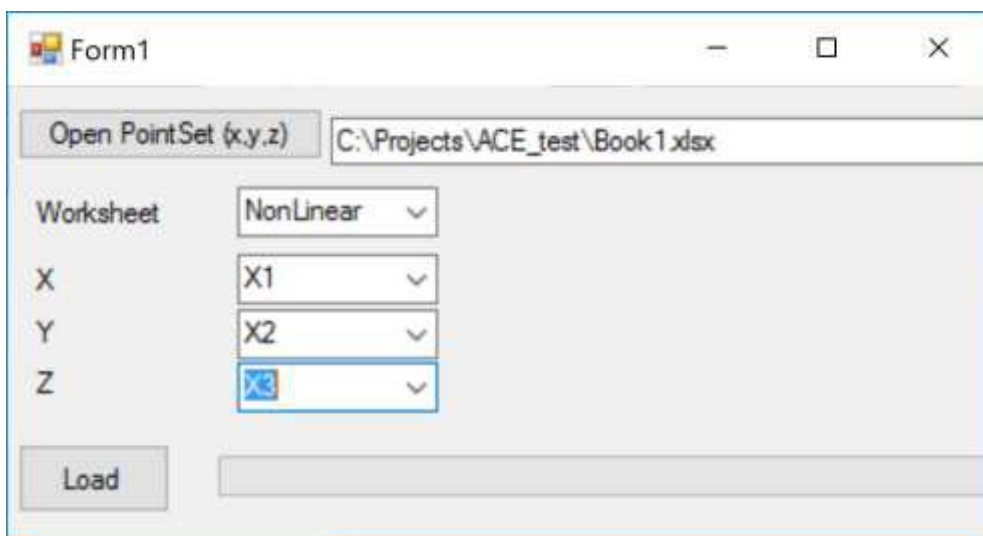
$$\varphi^*_3(X3) = X3^2$$

$$\varphi^*_4(X4) = X4^3$$

$$\varphi^*_5(X5) = X5$$

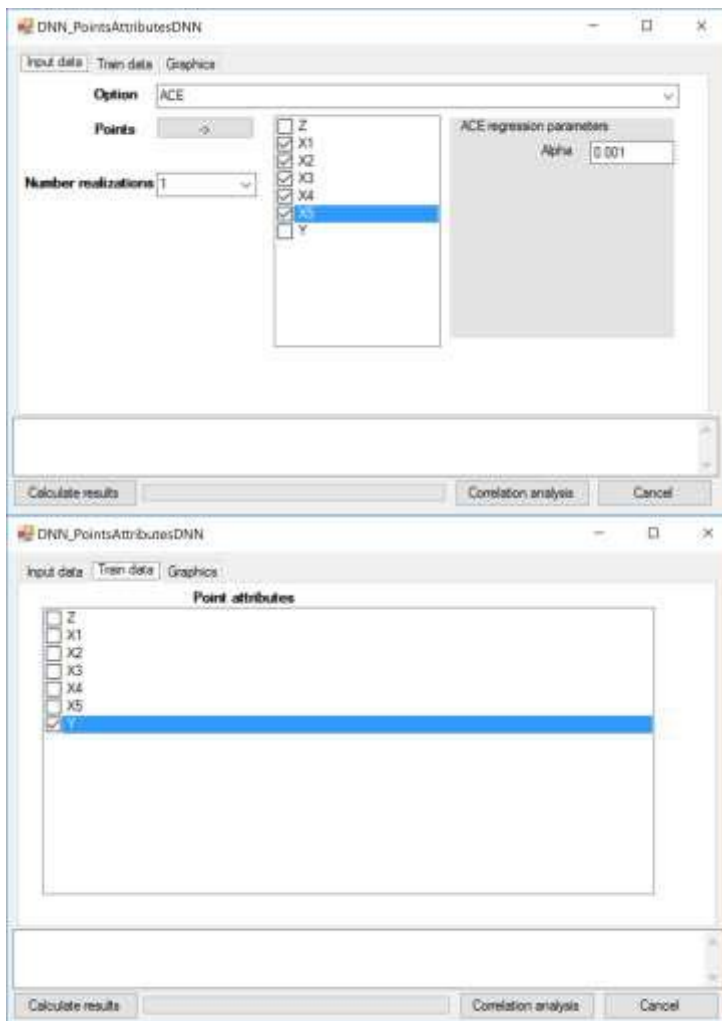
Total need to create more than 100 rows

2. Load the Excel worksheet using Excel Points attribute loader:



The 'Form1' dialog box is used to load an Excel worksheet. It features a text field for the file path, currently set to 'C:\Projects\ACE_test\Book1.xlsx'. Below this, there are dropdown menus for 'Worksheet' (set to 'NonLinear'), 'X' (set to 'X1'), 'Y' (set to 'X2'), and 'Z' (set to 'X3'). A 'Load' button is located at the bottom left.

3. Do ACE regression analysis using Points Property Prediction module:



The 'DNN_PointsAttributesDNN' dialog box is used for ACE regression analysis. It has three tabs: 'Input data', 'Train data', and 'Graphics'. The 'Input data' tab is active, showing 'Option' set to 'ACE' and 'Points' set to 'Z'. A list of variables (Z, X1, X2, X3, X4, X5, Y) is shown, with 'X3' selected. The 'Number realizations' is set to '1'. The 'ACE regression parameters' section shows 'Alpha' set to '0.001'. At the bottom, there are buttons for 'Calculate results', 'Correlation analysis', and 'Cancel'.

The 'DNN_PointsAttributesDNN' dialog box is also shown with the 'Train data' tab active. It displays a list of 'Point attributes' (Z, X1, X2, X3, X4, X5, Y) with 'Y' selected. The same buttons are present at the bottom.

- See results using Function window- all transform function can be restored by ACE algorithm:

