Sequential Piecewise Multiple Linear Regression Software Program Version 1.1.

User Guide

Nhat-Duc Hoang, Ph.D.

Faculty of Civil Engineering, Institute of Research and Development, Duy Tan University

Email: hoangnhatduc@dtu.edu.vn

This report describes the sequential piecewise multiple linear regression (SPMLR) [1] as a tool for performing nonlinear function approximation. A SPMLR software program has been developed and compiled in MATLAB.

Multiple linear regression (MLR) is commonly used data modeling tool due to its transparent model structure. However, MLR cannot model nonlinear data accurately. To preserve the advantage of the MLR and to enhance its capability when dealing with nonlinearity, researchers and practitioners have resorted piecewise multiple linear regression (PMLR) [2] [3].

A PMLR is a data modeling tool that uses each linear model to fit a subset of the input data *X*. The transition location from one explanatory variable domain to another explanatory domain is called as a breakpoint or a knot. A knot refers to the value of the explanatory variable where slopes of linear functions change [4].

The knot locations divide the overall learning into subspaces within which multiple linear models can be trained to approximate the local functional mapping. Notably, the appropriate value of a knot is not known and must be identified from the data set. A simple sequential algorithm is described in [1], relied on the concept of a hinge function [5], to automatically construct the PMLR based on the collected data set. The program GUI is provided in **Fig. 1.**

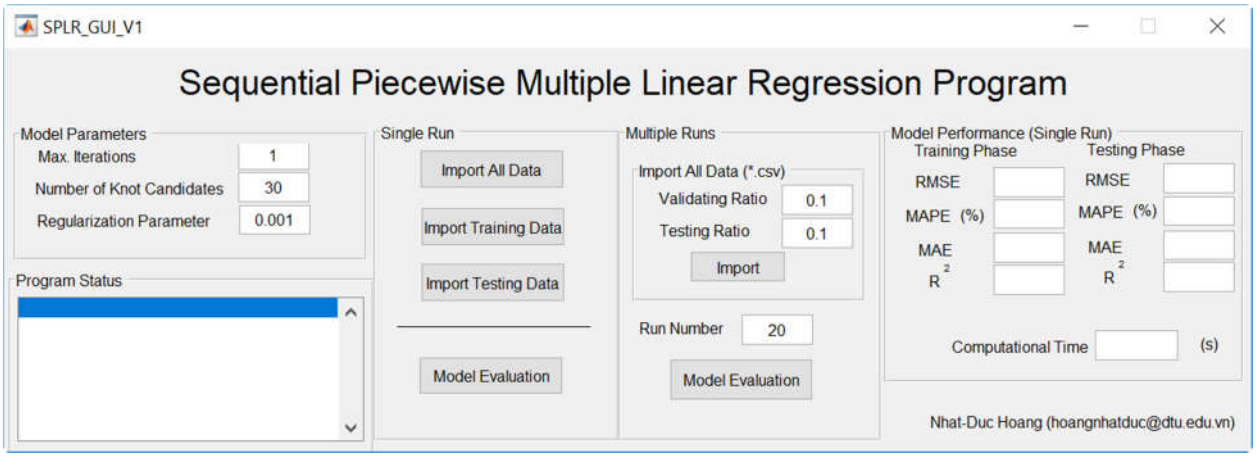

Fig. 1 SPMLR program

The SPMLR program supports both single run of model training and testing and multiple runs for model evaluation purpose. The model parameters include the maximum number of iterations, the number of knot candidates, and the regularization parameter.

For a single run, the user needs to provide the whole data set, the divided training set, and the divided testing set. The format of the dataset is .csv. The program will automatically normalize the dataset using Zscore normalization, perform model training, and perform prediction with data converted to the original range. The separated data sets are stored in the folder "DividedDataSets". The model results are provided in the folder "ModelPredictionResult". The file of ModelTestPerformance.csv stores the model performance of the testing phase (see **Fig. 2)**.

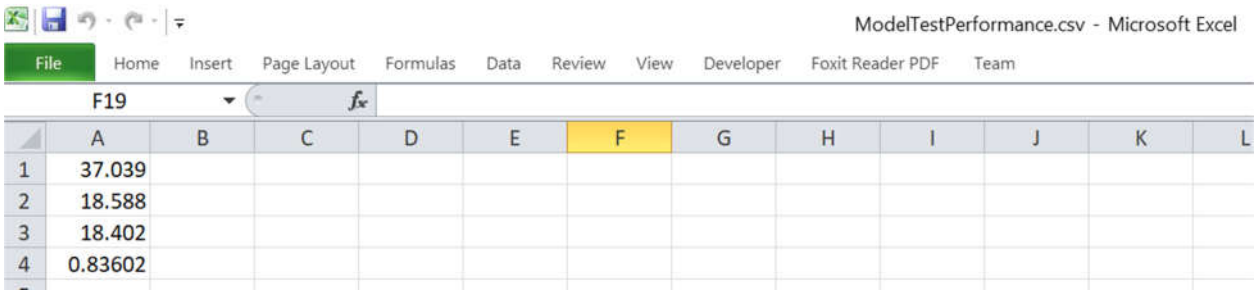

Fig. 2 The model performance of the testing phase (RMSE, MAPE, MAE, and R^2, respectively)

For multiple runs, the model performance is stored in the files of **MeanModelTestPerformance**, **MeanModelTrainPerformance**, **stdModelTestPerformance**, **stdModelTrainPerformance** which show the average and the standardeviation of the outcomes (RMSE, MAPE, MAE, and R^2).

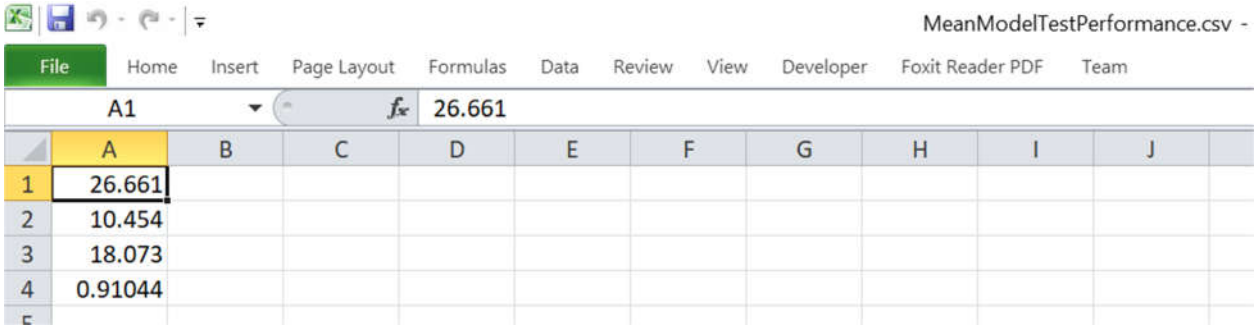

Fig. 3 The average model performance of the testing phase (RMSE, MAPE, MAE, and R^2, respectively)

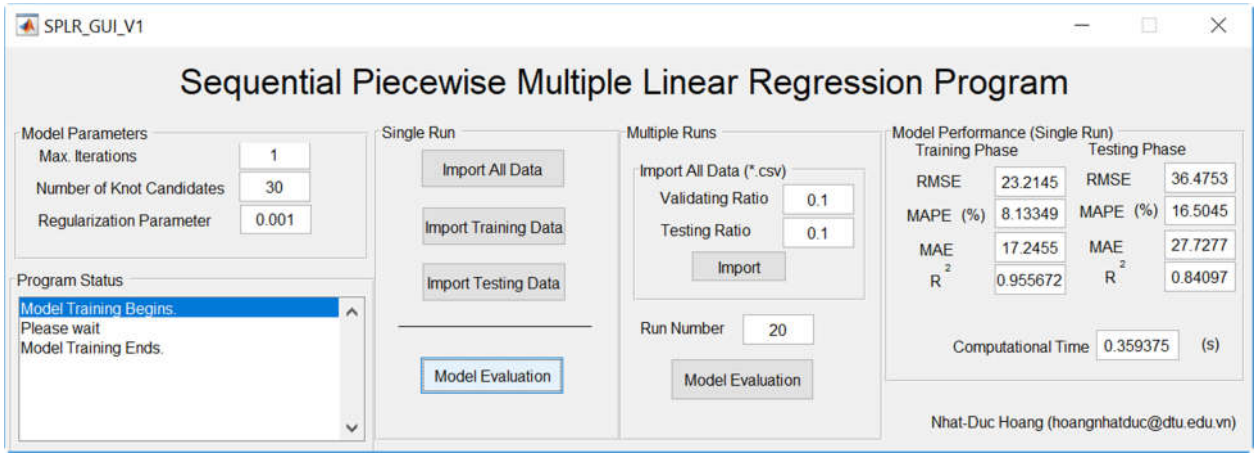

Fig. 4 The program training and prediction phases

The program prediction performance is shown in the section of Model Performance (**Fig. 4**). In case of multiple runs, the mean of outcomes is displayed.

The model structure is provided in the folder **ModelPredictionResult/ ModelStructure** as follows:

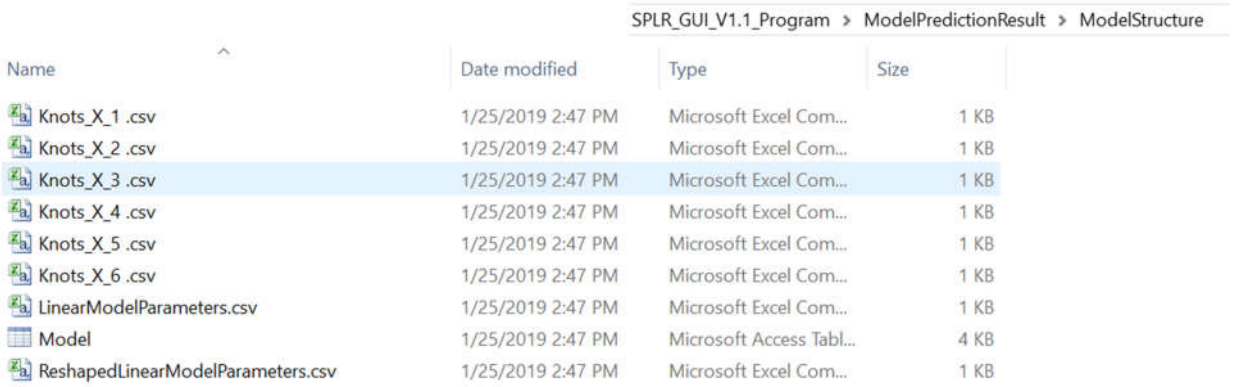

Fig. 5 Files that store the model structure

The files of Knots_X_i stores the knot values of the i^{th} variable. The file of **ReshapedLinearModelParameters.csv** stores the parameters of the linear model as described in [1]. Each column of the file includes 5 elements which are the parameters of the linear model.

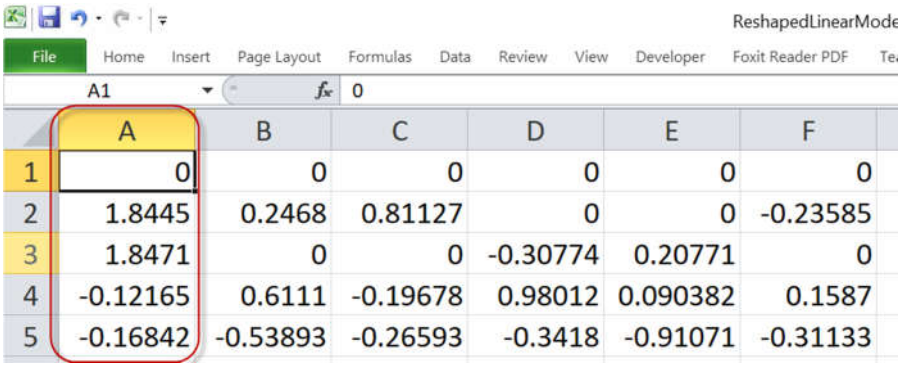

Fig. 6 File that store the parameters of the linear models

For instance, if each variable has 1 knot, then the first column is the parameters of the linear model of the 1^{st} variable; the second column is the parameters of the linear model of the 2^{nd} variable and so on.

However, if the variable X1 has 2 knots, then the first column is the parameters of the linear model of the 1^{st} knot of X1. The second column is the parameters of the linear model of the 2^{nd}

knot of X1. Hence, the third column is the parameters of the linear model of the 1st knot of X2, and so on.

The program can be downloaded at:

https://drive.google.com/open?id=1NKDaYsioWFonRfqq5Rb6611W9eqmz1Ip

Note: To run the program, the user needs to download and install MATLAB Runtime version R2017b (9.3) at:

https://www.mathworks.com/products/compiler/matlab-runtime.html

Release (MATLAB Runtime Version#)	Windows	Linux	Mac
R2018b (9.5)	64-bit	64-bit	Intel 64-bit
R2018a (9.4)	64-bit	64-bit	Intel 64-bit
R2017b (9.3)	64-bit ←	64-bit	Intel 64-bit

Reference

1. Hoang N-D (2019) Estimating Punching Shear Capacity of Steel Fibre Reinforced Concrete Slabs Using Sequential Piecewise Multiple Linear Regression and Artificial Neural Network. Measurement. doi:https://doi.org/10.1016/j.measurement.2019.01.035
2. Papadimitriou E, Mylona V, Golias J (2010) Perceived Level of Service, Driver, and Traffic Characteristics: Piecewise Linear Model. Journal of Transportation Engineering 136 (10):887-894. doi:doi:10.1061/(ASCE)TE.1943-5436.0000154
3. Greene ME, Rolfson O, Garellick G, Gordon M, Nemes S (2015) Improved statistical analysis of pre- and post-treatment patient-reported outcome measures (PROMs): the applicability of piecewise linear regression splines. Quality of Life Research 24 (3):567-573. doi:10.1007/s11136-014-0808-3
4. Ryan SE, Porth LS (2007) A tutorial on the piecewise regression approach applied to bedload transport data. Gen Tech Rep RMRS-GTR-189 Fort Collins, CO: US Department of Agriculture, Forest Service, Rocky Mountain Research Station 41 p. doi:https://doi.org/10.2737/RMRS-GTR-189
5. Breiman L (1993) Hinging hyperplanes for regression, classification, and function approximation. IEEE Transactions on Information Theory 39 (3):999-1013. doi:10.1109/18.256506