README

This README contains brief descriptions of the codes in this folder, as well information about where to access relevant data files.

Note that all python codes are written using Python 2.7 and Keras (https://keras.io/) with Theano as a backend. As Theano is currently being phased out of Keras, future projects will use Python 3 and Keras with TensorFlow.

CODES

BuildTrainingTestingData.py: Script to build the training and testing data sets. When run, creates the new testing/training data sets "Testing_Tmp.h5" and "Training_Tmp.h5."

TrainModel.py: Script to read in training data and train the neural network.

modelfunctions.py: Assorted functions for reading and writing CSV and HDF files, creating and training the neural network, etc.

EvalModel.py: Evaluate the trained network on the testing data.

Calculate_ROC_and_AUC_NN.py: Script to calculate and save (for later plotting) the ROC curves and associated AUC values associated with the neural network model for each slip distribution in the testing data set. Saves these quantities to "AllROC_NN.pkl."

Calculate_ROC_and_AUC_OtherQuantities.py: Script to calculate and save (for later plotting) the ROC curves and associated AUC values associated with maximum shear stress change, von-Mises criteria change, and Coulomb failure stress change for each slip distribution in the testing data set. Saves these quantities to "AllROC_OtherQuantities.pkl."

PlotThreeTestCases.py: Script to plot three cases from the testing data set.

Plot_ROC_Curves.py: Using the pickle files generated by *Calculate_ROC_and_AUC_NN.py* and *Calculate_ROC_and_AUC_OtherQuantities.py*, this script plots ROC curves for maximum shear stress change, the von-Mises yield criterion change, Coulomb failure stress change, and the NN.

GenerateInputs_IdealizedOkadaCase.m: This MATLAB script generates the input features (the components of the stress change tensor) for an idealized strike-slip earthquake case (note this script requires Brendan Meade's triangular dislocation code available at: https://github.com/brendanjmeade/tde).

RunNNPrediction_IdealizedOkadaCase.py: This script runs the neural network forward prediction of the aftershock probabilities for the idealized strike-slip earthquake case, using the input files generated by *GenerateInputs_IdealizedOkadaCase.m*.

PlotQuantitiesIdealizedFault.m: This MATLAB script plots maximum shear stress change, von-Mises criteria change, Coulomb failure stress change, and the NN forecast for an idealized strike-slip case.

DATA FILES
(available at: https://drive.google.com/drive/folders/1lAHfdjFd-Uv3wJcA0Tk2ViDIZeFt0mCA?usp=sharing)

AllCSV: This folder contains CSV files for each slip distribution used in this project. Each file has the following columns:
x: x location of grid cell centroid (UTM)
y: y location of grid cell centroid (UTM)

z: z location of grid cell centroid (UTM)
stresses_full_xx: xx component of stress change tensor at location
stresses_full_xy: xx component of stress change tensor at location
stresses_full_yy: yy component of stress change tensor at location
stresses_full_xz: xz component of stress change tensor at location
stresses_full_yz: yz component of stress change tensor at location
stresses_full_zz: zz component of stress change tensor at location
stresses_full_max_shear: maximum shear stress change at location
stresses_full_cfs_1: Coulomb failure stress change at location (sign convention 1)
stresses_full_cfs_2: Coulomb failure stress change at location (sign convention 2)
stresses_full_cfs_3: Coulomb failure stress change at location (sign convention 3)
stresses_full_cfs_4: Coulomb failure stress change at location (sign convention 4)
von_mises: von-Mises stress criterion change at location
aftershocksyn: 0 if aftershock did not occur within one year of mainshock in this grid cell; 1 if an aftershock did occur

*Testing_Filenames.h5***:** This index HDF file contains one list called "file_names_testing" – it is a list of the slip distributions randomly selected for the testing data set.

Training_Filenames.h5: This index HDF file contains one list called "file_names_training" – it is a list of the slip distributions randomly selected for the training data set.

Training.h5: This HDF file contains the training data organized into 7 data sets: the six features (*stresses_full_xx*, *stresses_full_xy*, *stresses_full_yy*, *stresses_full_xz*, *stresses_full_yz*, *stresses_full_zz)* and corresponding labels (*aftershocksyn)*.

Testing.h5: This HDF file contains the testing data organized into 7 data sets: the six features (*stresses_full_xx*, *stresses_full_xy*, *stresses_full_yy*, *stresses_full_xz*, *stresses_full_yz*, *stresses_full_zz)* and corresponding labels (*aftershocksyn)*.

1999CHICHI01MAxx.h5, 1995KOBEJA01YOSH.h5, 2005KASHMI01SHAO.h5: HDF files containing the x, y (UTM) coordinates of the slip distributions used in PlotThreeTestCases.py

TheBestWeights.h5: An HDF file containing the weights/biases associated with the trained neural network.