User manual

An artificial neural network ensemble approach to generate air pollution maps

S. Van Roode, J.J. Ruiz-Aguilar, J. González-Enrique, I.J. Turias

The aim of this study is to estimate concentration values of any phenomenon at unsampled location in order to produce maps. An artificial neural network (ANN) ensemble is proposed. The ensemble is composed by outputs of two different methods such as inverse distance weight (IDW) and least absolute shrinkage and selection operator (LASSO).

In this example, the validation and test processes have not been developed but they must be performed in order to select the best model parameters.

Main steps

Step 1: Generate model and outputs of LASSO as inputs for the neural network.

Step 2: Generate model and outputs of IDW as inputs for the neural network. IDW is a deterministic model so it is not necessary to train the model.

Step 3: Generate the ensemble model. A shallow neural network is performed whose inputs are the outputs of the other previous models (step 1 and step 2).

Step 4: Once the model is constructed, any point on the map can be estimated.

Example database

coord.mat: Grid references where columns are *(x_coord,y_coord)* of each monitoring station and rows are the number of each monitoring station. See Figure (1).

x.mat: NO_2 database where columns represent the monitoring stations and rows represent the different points of time. Database to create the models. See Figure (2).

x_new.mat: NO_2 database where columns represent the monitoring stations and rows represent the different points of time. New database to test the models. See Figure (2).

Monitoring stations	X coord. (UTM)	Y coord. (UTM)
	1	2
1	279239	4001847
2	286237	4006469
3	280980	4007826
4	285698	4009196
5	276184	4007408
6	285021	4009758
7	280289	4004653
8	281205	4006069
9	281534	4010206
10	289371	4005695
11	285910	4007229
12	283147	4006841
13	288757	4004181
14	283811	4009303

Fig. 1.

t	Monitoring stations													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	26.8300	9	1.5000	36.8300	16.6700	7.3300	24.1700	20.8300	15.5000	31.8300	14.1700	10.6700	46.5000	10.6700
2	24	13.6700	3	31.6700	13.8300	14.5000	22.5000	16.1700	19.3300	38	26.5000	16.8300	59	10.8300
3	46.5000	13.8300	6.1700	53.6700	15.3300	28.8300	38.6700	28.5000	30.8300	60.6700	35.8300	29.8300	58.5000	25.1700
4	47.8300	13.8300	10.1700	51.1700	22.1700	37.5000	44	40.3300	39.5000	45.6700	34.3300	34.1700	32	35.1700
5	36	9.5000	11.1700	39.1700	30.6700	37.8300	23.5000	24.5000	46.8300	37.1700	21.5000	17.3300	52.1700	30
6	20.8300	14.5000	13.8300	53.8300	38.6700	38.6700	21.6700	29.6700	53.1700	40.6700	43.6700	31.5000	56.1700	25.5000

Fig. 2

Example output

An example output data is displayed as a result of an experiment. If the experiment is replayed, the result varies subtly due to the stochastic nature of some of the models implemented. The results are calculated for the following input parameters: *cv = 2; b = 2.4; nhiddens = 14; t = 1 and spaced = 200.*

Functions

Main_script.m

Code that pretends to reproduce the experiment in the most intuitive way for the user. First, you must generate all three models (LASSO, IDW and ANN ensemble). Then, you can estimate any point on the map or generate a map (grid). Finally, you can display the map with scaled colours.

f_lasso_mdl.m

This function generates a structure with the lasso model information.

Inputs	
x	Database is a matrix where columns are monitoring stations and rows are the different points of time. See the structure in Figure (2).
coord	Grid references where columns are *(x_coord,y_coord)* of each monitoring stations and rows are the different points of time. See the structure in Figure (1).
cv	k-fold cross validation.
Outputs	
mdl	Structure with the lasso model information.
Options	
'InteractionPar'	Add to the new database a variable type that is an interaction between the measured variable and the distance. It is false by default.
'Standarize'	It is true by default.
'Criteria'	Specifies the lambda value selection criteria. 'mimMSE' Lambda value with the minimum MSE. By default. '1SE' Largest Lambda value such that MSE is within one standard error of the minimum MSE.

f_lasso_predict.m

This function estimates the concentration values for any point with LASSO model.

Inputs	
x_new	New predictor input values of each new point, specified as a matrix, it must have the same number of variables (columns) as war used to create mdl. See the structure in Figure (2).
coord_new	New predictor input coordinates or grid references, specified as a matrix, it must have *(x_coord,y_coord)* of each new point. See the structure in Figure (1).
coord	Grid references where columns are *(x_coord,y_coord)* of each monitoring stations and rows are the different points of time. See the structure in Figure (1).
Outputs	
y_new	Predicted response values evaluated at *x_new*, returned as a numeric vector. It is the same size as *x_new*.
Options	

'InteractionPar'	Add to the new database a variable type that is an interaction between the measured variable and the distance. It is false by default.

f_lasso_input_mdl.m

This function creates a new database as input LASSO database. It is used internally by the *f_lasso_mdl.m* function.

Inputs	
x	Database is a matrix where columns are monitoring stations and rows are the different points of time. See the structure in Figure (2).
coord	Grid references where columns are *(x_coord,y_coord)* of each monitoring stations and rows are the different points of time. See the structure in Figure (1).
Outputs	
x_lasso_input_mdl	A new database as input of lasso model where rows are input variables for a monitoring stations and for a moment in time.
y_lasso_input_mdl	A new target variable of lasso model.
Options	
'InteractionPar'	Add to the new database a variable type that is an interaction between the measured variable and the distance. It is false by default.

f_idw_predict.m

This function estimates the concentration values for any point whit IDW model.

Inputs	
x_new	New predictor input values of each new point, specified as a matrix, it must have the same number of variables (columns) as war used to create mdl. See the structure in Figure (2).
coord_new	New predictor input coordinates or grid references, specified as a matrix, it must have *(x_coord,y_coord)* of each new point. See the structure in Figure (1).
coord	Grid references where columns are *(x_coord,y_coord)* of each monitoring stations and rows are the different points of time. See the structure in Figure (1).
b	It a parameter that is used to highlight the spatial relationship between points. Farther points will be less important larger *b*.
Outputs	
y_new	Predicted response values evaluated at *x_new*, returned as a numeric vector. It is the same size as *x_new*.

f_ann_mdl.m

This function generates a structure with the lasso ANN information.

Inputs	
x	Outputs of previous models as network input.
y	Network targets (expected values)
nhiddens	Number of units or hidden neurons.
Outputs	
net	Structure with the ANN model information.