

**SERVIR**  **HIMALAYA**



**ICIMOD**

**Developing Dynamic Web-GIS Based Early  
Warning System for the Communities at  
Landslide Risks in Chittagong Metropolitan  
Area, Bangladesh**

**JUNE, 2015**



BUET-Japan Institute of Disaster Prevention and Urban Safety  
Bangladesh University of Engineering and Technology (BUET),  
Dhaka-1000, Bangladesh.

*Email: [jidpusinfo@jidpus.buet.ac.bd](mailto:jidpusinfo@jidpus.buet.ac.bd), Phone: +8802-9662975, Fax: +8802-9662975*

**Report on  
Web-GIS based Early Warning System Development**

**Submitted to**

International Centre for Integrated Mountain Development (ICIMOD),  
Kathmandu, Nepal

**Submitted by**

BUET-Japan Institute of Disaster Prevention and Urban Safety (BUET-  
JIDPUS); Bangladesh University of Engineering and Technology (BUET),  
Dhaka-1000, Bangladesh

**Authors**

**Bayes Ahmed**

Institute for Risk and Disaster Reduction,  
Department of Earth Sciences,  
University College London (UCL), UK  
Email: [bayes.ahmed.13@ucl.ac.uk](mailto:bayes.ahmed.13@ucl.ac.uk)

&

**Carlos Andres Osorio Murillo, PhD**

Postdoctoral Researcher,  
University of California, Berkeley, USA  
Email: [carosocali@gmail.com](mailto:carosocali@gmail.com)

June 2015



## Developing Dynamic Web-GIS Based Early Warning System for the Communities at landslide Risks in Chittagong Metropolitan Area, Bangladesh

---

### MULTINOMIAL LOGISTIC REGRESSION

To produce the susceptibility map of CDA for developing the Web-GIS based early warning system, MULTILOGISTICREG modelling method was chosen. MULTILOGISTICREG performs a multinomial logistic regression from image or attribute values files in which the input dependent variable is multi-categorical in nature and where there are continuous independent variables. Multinomial logistic regression is an extension of the conventional binomial logistic regression and employs the maximum likelihood estimation (MLE) technique.

In IDRISI Taiga software, the MULTILOGISTICREG performs a multinomial logistic regression in which the input dependent variable must be multinomial or multi-categorical in nature. That is, it must contain integer numbers in the range of -32767 to 32767, and in any sequence. It is assumed that the dependent variable is nominal.

Multinomial logistic regression is an extension of the conventional binomial or dichotomous logistic regression. Typically, more than two response categories are considered simultaneously in order to describe the composite relationship between one or more continuous independent variable(s) to the dependent variable.

Multinomial Logistic Regression (MLR) shares the same basic logic as Binomial Logistic Regression (BLR). Both assume the dependent variable following the logistic probability distribution and use the Maximum Likelihood Estimation as the estimation method. For a dependent variable with K response categories, MLR requires K-1 logit equations, one for each category relative to the reference category, to describe the relationship between the dependent variable and the independent variables. Mathematically, it is represented by the following functions, given category 0 as the reference category:

$$g_1(x) = \ln \left[ \frac{P(Y = 1 | X)}{P(Y = 0 | X)} \right] = \beta_{10} + \beta_{11}x_1 + \beta_{12}x_2 + \dots + \beta_{1q}x_q = X'\beta$$

$$g_2(x) = \ln \left[ \frac{P(Y = 2 | X)}{P(Y = 0 | X)} \right] = \beta_{20} + \beta_{21}x_1 + \beta_{22}x_2 + \dots + \beta_{2q}x_q = X'\beta$$

and

$$g_k(x) = \ln \left[ \frac{P(Y = k | X)}{P(Y = 0 | X)} \right] = \beta_{k0} + \beta_{k1}x_1 + \beta_{k2}x_2 + \dots + \beta_{kq}x_q = X'\beta$$

$g_k(x)$  is the logit function of category  $k$  against the reference category,

$X$  is the independent variable (covariate) vector,  $X = (x_1, x_2, \dots, x_q)'$ ;

$\beta_k$  is the coefficient vector estimated for category  $k$ ,  $\beta_k = (\beta_{k0}, \beta_{k1}, \dots, \beta_{kq})'$ .

Thus, given the independent variable vector  $X$ , the conditional probabilities of each outcome category can be calculated by the following formula:

$$P(Y = 0 | X) = \frac{1}{1 + e^{g_1(X)} + e^{g_2(X)} + \dots + e^{g_k(X)}}$$

$$P(Y = 1 | X) = \frac{e^{g_1(X)}}{1 + e^{g_1(X)} + e^{g_2(X)} + \dots + e^{g_k(X)}}$$

and

$$P(Y = k | X) = \frac{e^{g_k(X)}}{1 + e^{g_1(X)} + e^{g_2(X)} + \dots + e^{g_k(X)}}$$

Like the binary logistic regression, MULTILOGISTICREG employs the Maximum Likelihood Estimation (MLE) method to find the best fit sets of parameters (coefficients).

The conjoint likelihood function needed to maximize takes the general form as follows:

$$l(\beta) = \prod_{i=1}^N \prod_{k=0}^K [\pi_k(X_i)^{y_{ik}}]$$

where:

$l(\beta)$  is the conjoint likelihood to be maximized;

$\pi_k(X_i)$  is a simplified notation for  $P(Y = k | X)$ ;

$y_{ik}$  is the observed value for category  $k$  at observation  $i$ , which equals to 1 if observed, otherwise 0, and thus

$$\sum y_{ij} = 1 \quad (j = 1, 2, \dots, k).$$

$N$  is the total number of observations.

To maximize the above likelihood function, it requires a solution for a group of nonlinear equations taking the following general form:

$$\frac{\partial L(\beta)}{\partial \beta_{jp}} = \sum_{i=1}^N x_{yi} (y_{ij} - \pi_{ij}) = 0$$

where:

$L(\beta)$  is the logarithm function of  $l(\beta)$  above;

$\partial L(\beta) / \partial \beta_{jp}$  is the first partial derivative of  $L(\beta)$  against the coefficient  $\beta_{jp}$  for  $j = 1, 2, \dots, k$  and  $p = 0, 1, 2, \dots, q$  with  $x_{0i} = 1$ ;

$\pi_{ij}$  is a simplified notation of  $\pi_j(X_i)$  for  $(j = 1, 2, \dots, k)$ .

In solving the above non-linear equations, MULTILOGISTICREG uses a variant of the Newton-Raphson algorithm.

### **MULTILOGISTICREG assumptions:**

- a. The dependent variable  $Y$  is assumed to be a random variable following the conditional multinomial logistic distribution. For any specific observation, the value of the dependent variable as category  $k$  is viewed as either being 1 if observed or 0 if not observed.
- b. The outcome on  $Y$  are assumed to be mutually exclusive and exhaustive.
- c.  $Y$  is assumed to depend on  $q$  observable explanatory variables and the relationship follows the non-linear logistic curve.
- d. The analytical data are generated by sampling with sample size  $N$ , sample points indexed by  $i$  ( $i = 1, 2, \dots, N$ ).

- e. There are no restrictions on the independent variables except they should not be linearly correlated and that N should be greater than the product of K-1 and q, i.e.,  $N > (K-1)q$ .
- f. The error term of each observation is assumed to be independent of the error terms of all other observations.

### **Output statistics:**

Like in binomial logistic regression, multinomial logistic regression uses the likelihood ratio principle to evaluate the goodness of fit of the model. The ratio is based on two statistics:

$$-2\text{Log}(L_0)$$

where  $L_0$  is the value of the likelihood function if all coefficients except the intercept are 0. This intercept-only model is actually the null model in question.

$$-2\text{Log}(L_m)$$

where  $L_m$  is the value of the likelihood function for the full model as fitted with the explanatory variables.

Logically, the log likelihood above is analogous to the Sum of Squared Errors (SSE) in the ordinary least square estimation (and thus for the null model, this is analogous to the total variance in data). The smaller the log likelihood of the model, the better the model fits the data. Based on these two statistics, other statistics for fitness measuring include:

$$\text{Model Chi Square: } \chi_m^2 = -2(\text{Log}(L_m) - \text{Log}(L_0))$$

This is also known as the likelihood ratio statistic which follows, approximately, a chi-square distribution when the null hypothesis is true. This statistic tests the hypothesis that all coefficients except the intercept are 0. Thus, it is a similar test as the F statistic in linear regression analysis. In the multinomial context, the degrees of freedom for this chi-square statistic is equal to the product of K-1 and q. K is the number of response categories in the dependent variable and q is the number of the independent variables.

$$\text{McFadden Pseudo } R_M^2 = 1 - (\text{Log}(L_m) / \text{Log}(L_0))$$

A McFadden pseudo R-square of 1 indicates a perfect fit, whereas a pseudo R-square equal to 0 indicates no relationship. Clark and Hosking (1986) comment that a McFadden Pseudo R-square greater than 0.2 can be considered a relatively good fit.

$$\text{McFadden Pseudo } R_M^2 = [1 - (L_0 / L_M)^{2/N}] / [1 - (L_0)^{2/N}]$$

Similarly, this pseudo R-square has the range of 0 to 1.

Finally, MULTILOGISTICREG will evaluate the model by cross-tabulating the dependent variable (rows) and the hardened prediction (columns) derived from the group of class membership files based on the Maximum Likelihood rule. Also output are the relevant validation statistics such as accuracy percentage, chi-square, Cramer's V, and p-level.

### **References:**

Clark, W.A., and P.L. Hosking. 1986. Statistical Methods for Geographers (Chapter 13). New York: John Wiley & Sons.

Hosmer, D.W., and S. Lemeshow. 2000. Applied Logistic Regression (2nd Edition). John Wiley & Sons, Inc.

Menard, S. 2001. Applied Logistic Regression Analysis. Sage University Papers Series on Quantitative Applications in the Social Sciences, 07-106. Thousand Oaks, CA: Sage.

Press, W.H., Flannery, B.P., Teukolsky, S.A., and W.T. Vetterling, 1986. Numerical Recipes (Chapter 9). Cambridge: Cambridge University Press.

Press, W.H., Teukolsky, S.A., W.T. Vetterling, and Flannery, B.P. 1992. Numerical Recipes in C: The Art of Scientific Computing (2nd Edition). Cambridge: Cambridge University Press.

Eastman 2012. IDRISI Taiga Tutorial.

### **Results:**

Analytical Results of Multinomial Logistic Regression:

/\* Note: For a dependent variable with a number of categories K, it requires K - 1

/\* equations, one for each category relative to the reference category,

/\* to describe the relationship between the dependent variable and the independent

/\* variables. The following statistics are based on these equations.

Default reference category: Category 0



$$G_1(X) = \text{Logit}(\text{Cat 1 v.s. Cat 0})$$

$$= -13.9942 - 0.3241(\text{DRAIN\_FINAL}) + 15.7952(\text{ELEVATION\_FINAL}) - 0.3068(\text{LC\_FINAL}) + 1.2044(\text{NDVI\_FINAL}) + 0.1860(\text{PRECIPITATION}) - 1.3146(\text{ROAD\_FINAL}) - 0.2674(\text{SLOPE\_FINAL}) + 2.0202(\text{WATER\_FINAL})$$

$$G_2(X) = \text{Logit}(\text{Cat 2 v.s. Cat 0})$$

$$= -9.7481 - 1.7475(\text{DRAIN\_FINAL}) + 14.9538(\text{ELEVATION\_FINAL}) - 1.3740(\text{LC\_FINAL}) - 0.4435(\text{NDVI\_FINAL}) - 0.5006(\text{PRECIPITATION}) + 0.6944(\text{ROAD\_FINAL}) - 0.6255(\text{SLOPE\_FINAL}) + 2.7979(\text{WATER\_FINAL})$$

$$G_3(X) = \text{Logit}(\text{Cat 3 v.s. Cat 0})$$

$$= -15.0379 - 0.6892(\text{DRAIN\_FINAL}) + 16.0513(\text{ELEVATION\_FINAL}) - 1.0162(\text{LC\_FINAL}) + 1.2640(\text{NDVI\_FINAL}) + 0.6459(\text{PRECIPITATION}) - 2.1894(\text{ROAD\_FINAL}) - 0.0544(\text{SLOPE\_FINAL}) + 2.6101(\text{WATER\_FINAL})$$

$$G_4(X) = \text{Logit}(\text{Cat 4 v.s. Cat 0})$$

$$= -8.8974 - 0.3164(\text{DRAIN\_FINAL}) + 11.7855(\text{ELEVATION\_FINAL}) - 0.4808(\text{LC\_FINAL}) + 1.0340(\text{NDVI\_FINAL}) + 0.0977(\text{PRECIPITATION}) - 0.8818(\text{ROAD\_FINAL}) - 0.3705(\text{SLOPE\_FINAL}) + 1.9079(\text{WATER\_FINAL})$$

$$G_5(X) = \text{Logit}(\text{Cat 5 v.s. Cat 0})$$

$$= -11.8704 - 0.9488(\text{DRAIN\_FINAL}) + 15.1529(\text{ELEVATION\_FINAL}) - 0.1047(\text{LC\_FINAL}) + 0.6240(\text{NDVI\_FINAL}) - 0.5619(\text{PRECIPITATION}) - 1.4557(\text{ROAD\_FINAL}) - 0.0179(\text{SLOPE\_FINAL}) + 2.8437(\text{WATER\_FINAL})$$

### 1. Case Processing Summary

-----

Class # | Actual observations | Samples processed

-----

0	1249297	276541
1	297013	65890
2	22794	5083
3	66277	14781
4	237061	52501
5	140476	31353
Total	2012918	446149

-----





## 2. Model Fitting Information

---

Model	Sub-models	-2 Log Likelihood	Chi-Square	df	Sig.
Full Model	Intercept Only	1053995.5336	-	-	-
	Fitted	378143.8271	675851.7064	40	0.0000

---

## 3. Pseudo R-Square

McFadden | 0.6412  
 Magelkerke | 0.0739

## 4. Classification

| Prediction

---

Observed	Cat 0	Cat 1	Cat 2	Cat 3	Cat 4	Cat 5	% Correct
Cat 0	1237015	1235	181	53	11218	1247	98.89%
Cat 1	2883	134232	1120	7179	118009	33590	45.19%
Cat 2	403	1533	13729	779	2426	3926	60.23%
Cat 3	116	34032	243	14238	14787	2861	21.48%
Cat 4	4801	46831	1683	1026	163214	19506	68.85%
Cat 5	2700	21414	313	69	39775	76205	54.25%
Overall %	61.94%	11.88%	0.86 %	1.16 %	17.35%	6.82 %	81.34%

---

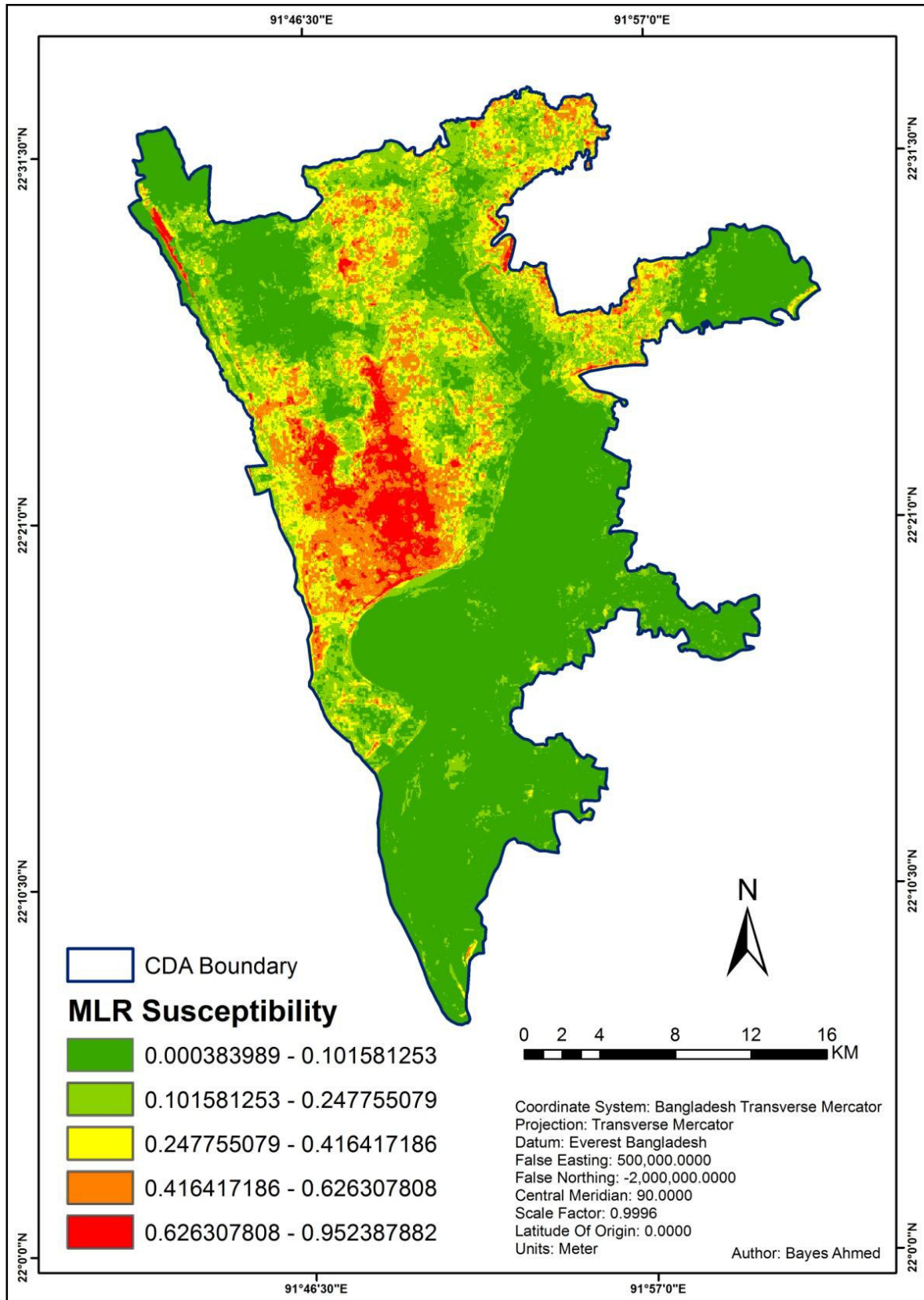
Chi Square = 3674485.3056

df = 25

P-Level = 0.0000; Cramer's V = 0.6040

The susceptibility map using MLR method is illustrated in Figure 1. For this analysis, the similar factor maps, as prepared in the landslide susceptibility report, were used.

Figure 1: Susceptibility map produced using MLR method.



### **Warning Classification:**

Now based on the rainfall pattern analysis results, the following 5 classes were created for developing the Web-GIS based landslide early warning system:

Susceptibility Class	Precipitation Range
0.000383989 - 0.101581253	> 250mm in 1 day
0.101581253 - 0.247755079	> 100mm in 1 day
0.247755079 - 0.416417186	> 40 mm in 1 day or > 200mm in continuous 3 days
0.416417186 - 0.626307808	> 25 mm in 1 day or > 50mm in continuous 3 days
0.626307808 - 0.952387882	>= 15mm in 1 day or > 30mm in continuous 3 days

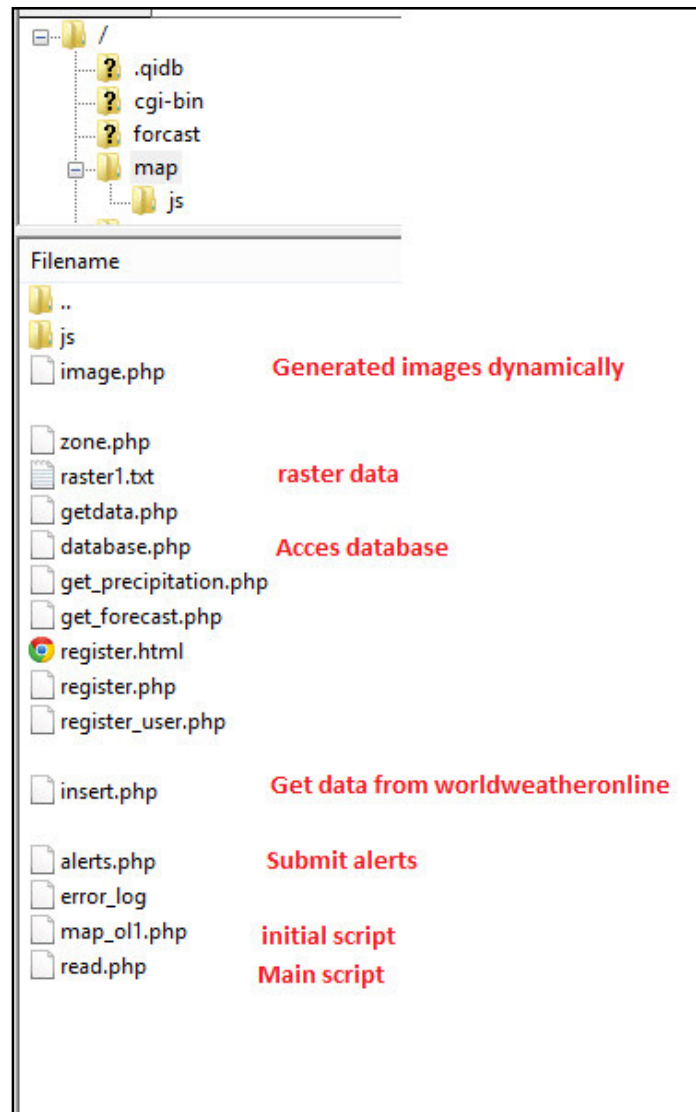
### **Developing the Web-GIS Early Warning System:**

At first, a website domain and hosting was registered for 3 years. Then the website was designed for developing the alert system. The alert system was developed combining the susceptible zones from susceptibility map and the rainfall threshold for corresponding zones. The Web Warning System (WWS) displays different risk areas according to precipitation in the study area. The precipitation data is updated from worldweatheronline.com once a day. The system is able to send warning emails to the registered users about the affected areas according to forecast precipitation. The WWS was developed using the following tools: PHP and javascript APIs: OpenLayes, W2UI and jQuery. The system receives a reclassified raster dataset to generate the risk areas. A special raster format was created to speed up the display of the risk areas (Annex- Raster format) using PHP. This new format is generated by a new plugin in dotspatial (<http://codeplex.dotspatial.com>) (Attached the c# solution and compiled version - Demomap.exe).

The precipitation and users data are stored in a MySQL database (Annex- Database schema). Images of the raster dataset are generated dynamically using PHP using the raster1.txt file (generated by dotspatial-plugin). In order to update the system, it is just needed to upload a new version of raster1.txt file.

The read.php file contains all PHPscripts that require WWS. The starting point of the application is the map\_ol1.php. The list of files require by WWS are shown in the figure below.

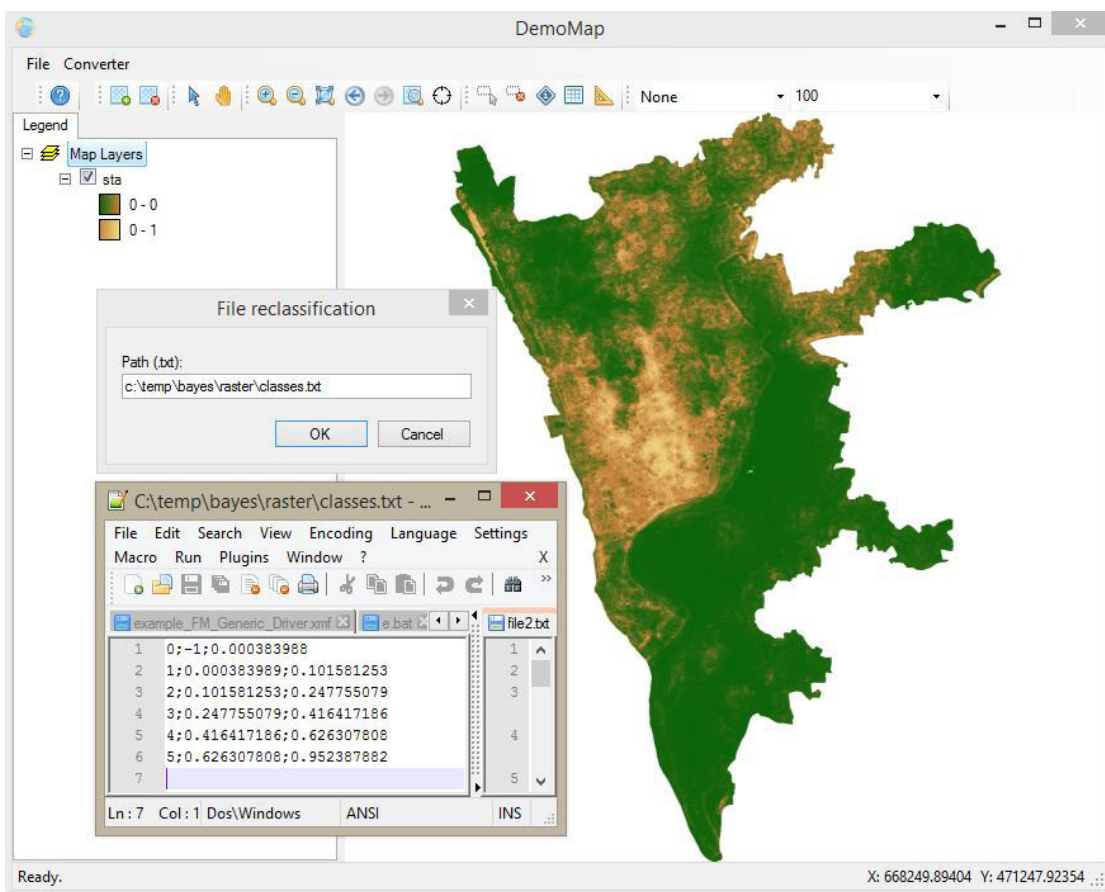
Figure 2: WWS files



## ANNEXURE

### 1. Reclassifying map using Dotspatial plugin

This plugin generates the reclassified file required by WWS. The user should provide a file with the classes of each range in a text file as is shown figure.

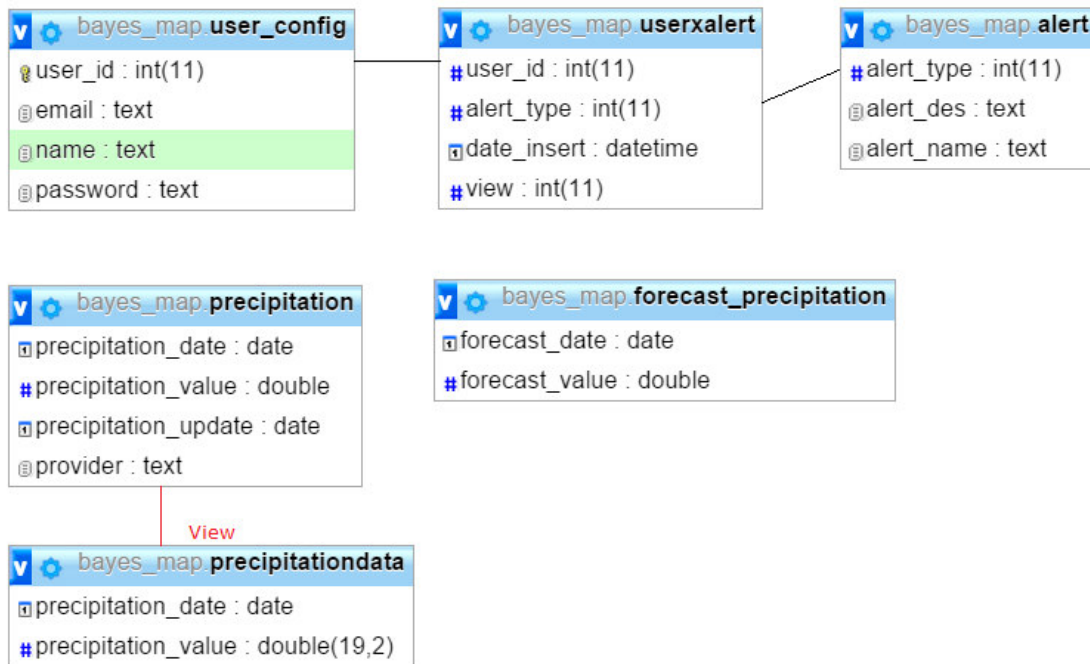


### 2. Raster Format

This format require a header with the following information minX ,maxY,nrows,n\_cols,cellsize,maxX,minY. The system assumes that the coordinate system is Bangladesh\_Transverse\_Mercator. For each row in the raster is generated a line with the class-number of cells of that class found. In the below example is found in the second line are found 701 cells of the class 0, 2 cells of the class 2 and finally 515 cells of the class 0.

673278.625 495512.1875 1654 1218 30 709818.625 445892.1875  
 #(xmin\_ymax\_nrows\_n\_cols\_cellsize\_maxX\_minY)  
 0-701 2-2 0-515  
 0-701 2-5 0-512  
 0-700 2-7 0-511  
 0-699 2-3 1-1 2-4 0-511  
 0-693 2-2 4-2 3-3 2-2 1-1 2-4 3-1 0-510  
 ...

### 3. Database schema



\*All the remaining codes and relevant files are attached in Codes folder.