2.8.1 SPATIAL DATA QUALITY
Mini Project Report

DIGITAL SOIL MAPPING FOR CHIRALA AREA USING MULTIPLE LINEAR REGRESSION

Date of submission: 11 March 2022

Submitted by:

DARIVEMULA PREETHAM
DEV DINESH
VIVEK KUMAR

M.Tech Geoinformatics

Indian Institute of Remote Sensing

Indian Space Research Organisation

Department of Space, Government of India

Dehradun-248001, Uttarakhand

March - 2022

Abstract:

Having an idea of soil quality and data about its nutrients is very important in the area of agriculture and in general so that the optimum use of land can be achieved, if the land is fit for cultivation it should be used for that, otherwise, it can be used for infrastructure and other purposes. If the information is available in the form of maps it helps to make the data, more consumable. The soil mapping technique is utilized for this purpose, traditional approaches relied heavily on the use of ground data but as the availability of satellite images and its use has gained more popularity Digital Soil Mapping (DSM) technique has emerged as a new game-changer, It is cost-effective and less time-consuming.

The phosphorous (P) and zinc (Zn) are important constituents of soil for agriculture and we aim to derive DSM of both for the Chirala block located in district Prakasham of Andhra Pradesh using regression techniques. We use features such as NDVI (Normalized Difference Vegetation Index), precipitation, and Soil surface temperature to establish a relationship with the two soil nutrients P and Zn, for which a statistical significance testing technique is used to obtain meaning full relation between the parameter and the feature.

Introduction:

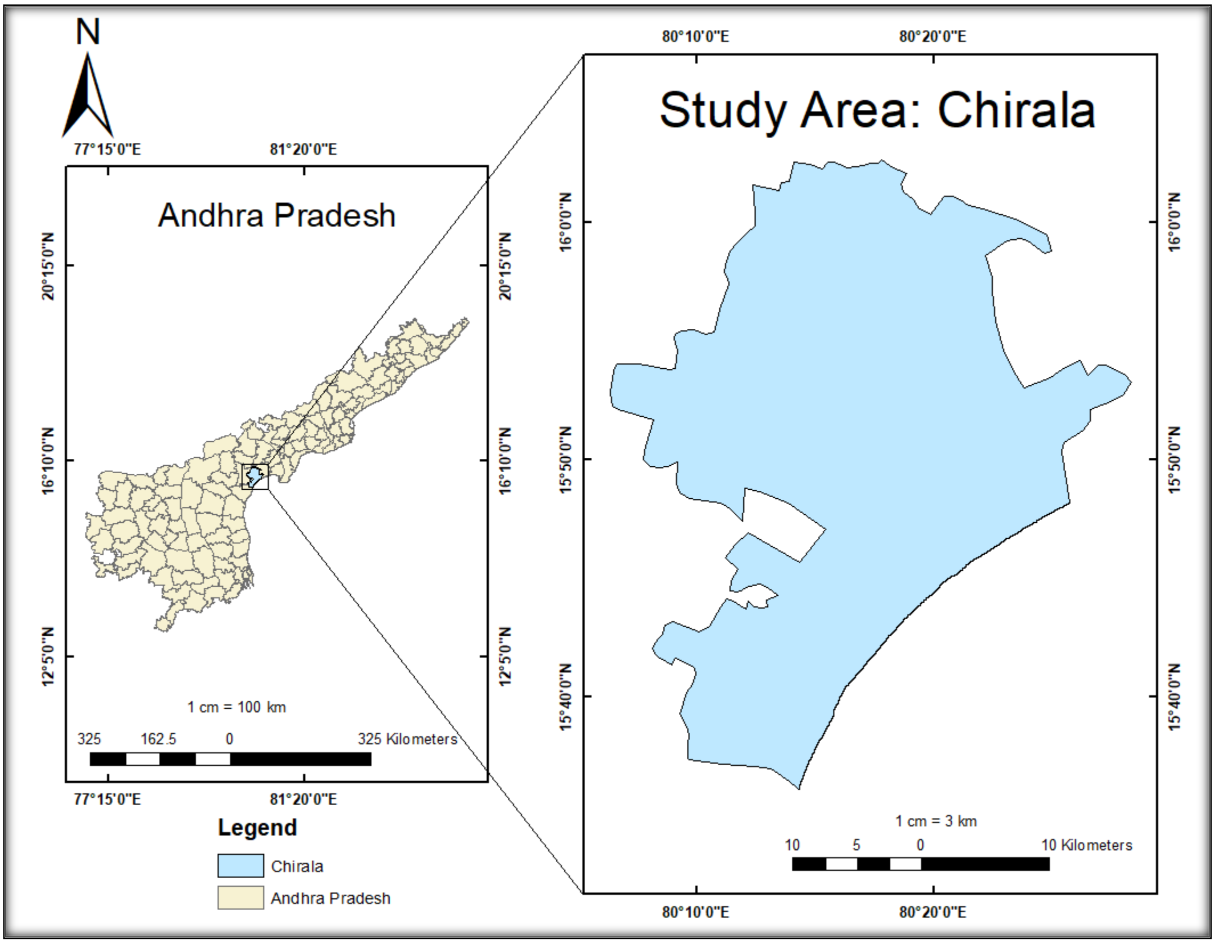
Digital soil mapping is a subfield in soil sciences, the technique got a major push post-2000.

Digital soil mapping is defined as the creation and population of spatial soil information systems by the use of field and laboratory observational methods coupled with spatial and non-spatial soil inference systems [2]

The increase in processing power of computers and subsequent availability of spatial data like terrain information and satellite images, advancement in machine learning techniques, and GIS software made it easier to quickly collect geotagged data rather than relying on traditional methods going on the ground and collecting GPS tagged samples. The clubbing of statistical techniques and spatial data processing made the work quick and highly cost-effective.

Andhra Pradesh is the bejewelled rice bowl of India and majorly depends on agriculture as 63% of the population is located in rural areas and depends heavily on agricultural produce for their livelihood. Prakasam District was formed in

1970 with the amalgamation of backward areas of erstwhile Guntur, Nellore, and Kurnool Districts with Headquarters at Ongole. Geographically located at the coastal regions of Bay of Bengal and at located at GPS coordinates 15° 48' 43.4664" N and 80° 21' 19.3572" E.



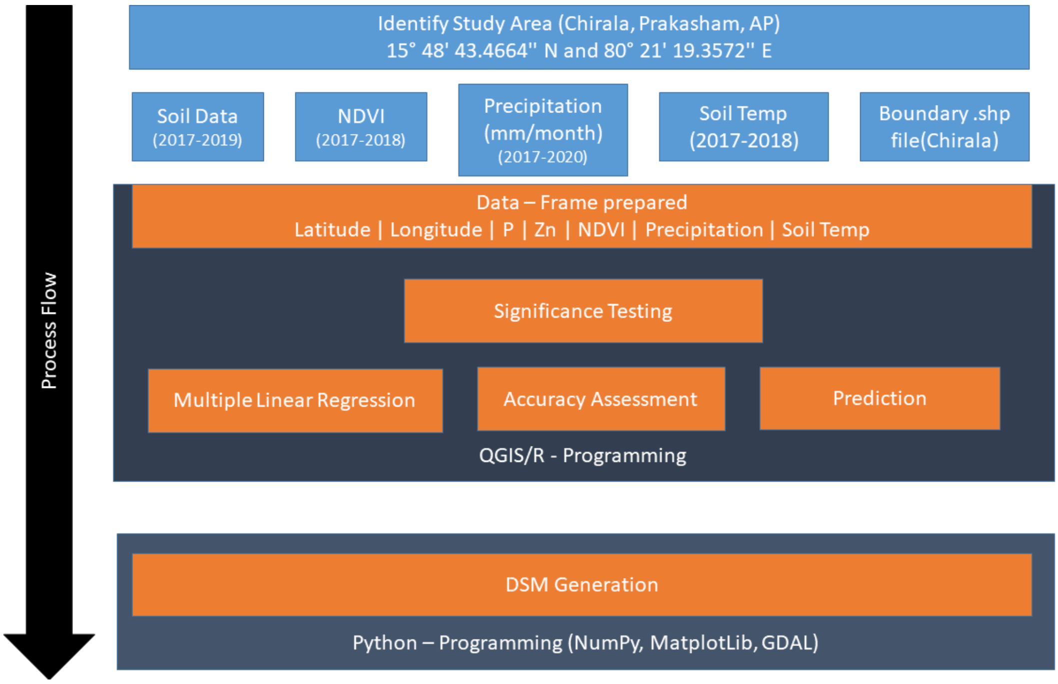
Phosphorus is one of the major plant nutrients in the soil. It is a constituent of plant cells, essential for cell division and the development of the growing tip of the plant. For this reason, it is vital for seedlings and young plants.

Zinc is an important component of various enzymes that are responsible for driving many metabolic reactions in all crops. Growth and development would stop if specific enzymes were not present in plant tissue. Carbohydrate, protein and chlorophyll formation is significantly reduced in zinc-deficient plants. Therefore, a constant and continuous supply of zinc is needed for optimum growth and maximum yield.

The use of MLR/ SLR (Multiple Linear Regression/ Simple Linear Regression) based on the significance test results and relation hence achieved after it will help us predict

Data and Methods.

DATA	Definition	Link
Soil Health	Geocoded – Phosphorous, and Zinc.	Soil Health Card (dac.gov.in)
Surface Temperature	GeoTIFF	Google Earth
NDVI	GeoTIFF	Google Earth
Precipitation	GeoTIFF	Geovani



Methodology:

The workflow of the process is stated above. The methodology used in this study is regression analysis and mainly the multiple variable linear regression approach is considered for this study. The variables considered for this study are NDVI (I year averaged data), Precipitation data (2 Year data, averaged on mm/month basis.), Soil temperature (are the predictor variables or the independent variables which predict the variables P and Zn separately. Here, Phosphorous(P) and zinc(Zn) are the dependent variables in the linear regression model. The independent variables are tested for significance and the model is updated for every iteration giving the new coefficients associated with the variables. In general, the linear regression model is given by,

$$Y = b0 + b1 * x$$

where,

Y is the dependent variable,

bo is the intercept coefficient and,
b1 is the coefficient of the slope.

The linear regression equation used in for this study is,

1.) For Phosphorous nutrient parameter:

$$P = b0 + b1 * NDVI + b2 * Precipitation + b3 * Soil\_temp$$

Where,

P is the phosphorous,
bo is the intercept,
b1 is the coefficient of the variable NDVI,
b2 is the coefficient of the variable Precipitation,
b3 is the coefficient of the variable Soil_temp,

2.) For Zinc nutrient parameter:

$$Zn = bo + b1 * NDVI + b2 * Precipitation + b3 * Soil\_temp$$

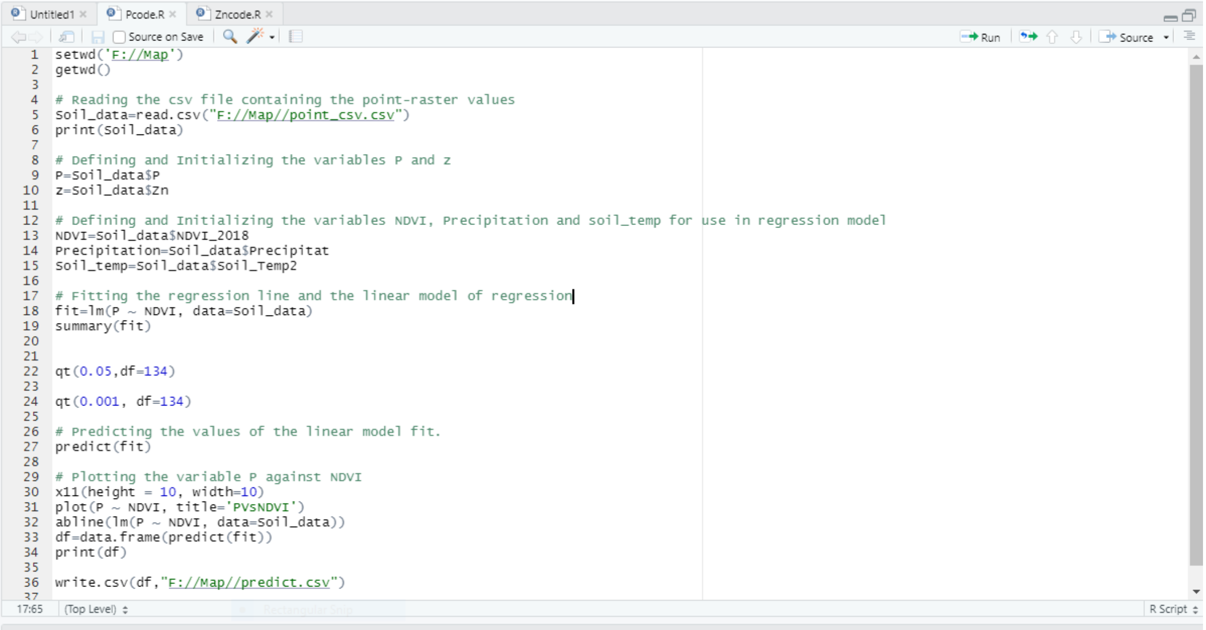
Where,

Zn= The parameter Zinc,
b0 is the intercept,
b1 is the coefficient of the variable NDVI,
b2 is the coefficient of the variable Precipitation,
b3 is the coefficient of the variable Soil_temp.

Multiple linear regression using R:

The regression analysis is carried out in the R software, and the testing for the significance of the variables is done in the same environment.

1.) For the phosphorous parameter the code in R is shown below,



```r
setwd('F://Map')
getwd()

# Reading the csv file containing the point-raster values
Soil_data=read.csv("F://Map//point_csv.csv")
print(Soil_data)

# Defining and Initializing the variables P and z
P=Soil_data$P
z=Soil_data$Zn

# Defining and Initializing the variables NDVI, Precipitation and soil_temp for use in regression model
NDVI=Soil_data$NDVI_2018
Precipitation=Soil_data$Precipitat
Soil_temp=Soil_data$Soil_Temp2

# Fitting the regression line and the linear model of regression
fit=lm(P ~ NDVI, data=Soil_data)
summary(fit)


qt(0.05,df=134)

qt(0.001, df=134)

# Predicting the values of the linear model fit.
predict(fit)

# Plotting the variable P against NDVI
x11(height = 10, width=10)
plot(P ~ NDVI, title='PVsNDVI')
abline(lm(P ~ NDVI, data=Soil_data))
df=data.frame(predict(fit))
print(df)

write.csv(df,"F://Map//predict.csv")
```

```
Summary of the regression model:

Call:
lm(formula = P ~ NDVI + Precipitation + Soil_temp, data = Soil_data)

Residuals:
    Min      1Q  Median      3Q     Max
-23.982 -11.271  -2.803   8.751  46.971

Coefficients:
               Estimate  Std. Error  t value  Pr(>|t|)
(Intercept)   1913.9270   2805.6035    0.682     0.496
NDVI           -14.7400     14.9758   -0.984     0.327
Precipitation    0.2073      0.4395    0.472     0.638
Soil_temp       -6.2967      9.2434   -0.681     0.497

Residual standard error: 15.5 on 134 degrees of freedom
Multiple R-squared:  0.01692,	Adjusted R-squared:  -0.005091
F-statistic: 0.7687 on 3 and 134 DF,  p-value: 0.5135
```

The summary results of the model are given by the R software and the testing is performed for the different coefficients of the independent variables at a significance level of 0.05.

By running the tests and for the iterations the statistically insignificant variables are excluded and the linear equation remains with the significant variable NDVI for Phosphorous(P).

```
The summary result of the above code is shown below:
Call:
lm(formula = P ~ NDVI, data = Soil_data)

Residuals:
   Min      1Q  Median      3Q     Max
-22.195 -11.030  -3.124   9.304  48.026

Coefficients:
             Estimate        Std. Error  t value       Pr(>|t|)
(Intercept)   25.000     3.642          6.864      2.16e-10 ***
NDVI        -16.129       14.859       -1.086      0.28
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


Residual standard error: 15.45 on 136 degrees of freedom
Multiple R-squared:  0.00859,    Adjusted R-squared:  0.0013
F-statistic: 1.178 on 1 and 136 DF,  p-value: 0.2796
```

For the phosphorous linear model, only the **NDVI** variable is found significant

2.) For Zinc parameter the code in R is shown below:

```r

# Reading the csv file containing the point_sample-raster values
Soil_data=read.csv("F://Map//point_csv.csv")
print(Soil_data)

# Defining and Initializing the variables P and z
P=Soil_data$P
z=Soil_data$Zn

# Defining and Initializing the variables NDVI, Precipitation and soil_temp for use in regression model
NDVI=Soil_data$NDVI_2018
Precipitation=Soil_data$Precipitat
Soil_temp=Soil_data$Soil_Temp2

#  Fitting the regression line and the linear model of regression
fit=lm(z ~ Soil_temp, data=Soil_data)
summary(fit)

# Plotting the graph between z and the soil_temp
x11(height=10, width=10)
y=z
x=Soil_temp
plot(y ~ x)
abline(lm(y ~x, data=Soil_data))

# Predicting the values of the independent variable for the given values of the dependent variables
predict(fit)
df1=data.frame(predict(fit))

write.csv(df1,"F://Map//Predictzn.csv")

```

```
Summary of the regression model:
Call:
lm(formula = z ~ NDVI + Precipitation + Soil_temp, data = Soil_data)

Residuals:
     Min       1Q   Median       3Q      Max
-0.71673 -0.26451 -0.05298  0.22901  2.41104

Coefficients:
                 Estimate   Std. Error   t value   Pr(>|t|)
(Intercept)    130.470049   74.420982    1.753     0.0819 .
NDVI            -0.208517    0.397244   -0.525     0.6005
Precipitation    0.003671    0.011658    0.315     0.7534
Soil_temp       -0.429776    0.245188   -1.753     0.0819 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4111 on 134 degrees of freedom
Multiple R-squared:  0.03453,    Adjusted R-squared:  0.01291
F-statistic: 1.597 on 3 and 134 DF,  p-value: 0.193
```

By running the significance test at significance level of **0.05** and for each iteration the insignificant variables are excluded and the variable found significant is the **soil temperature**.

```
The summary of the test for the above code:
Call:
lm(formula = z ~ Soil_temp, data = Soil_data)

Residuals:
    Min       1Q   Median       3Q      Max
-0.75277 -0.26313 -0.05274  0.21710  2.40386

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 142.3898    67.1862   2.119   0.0359 *
Soil_temp    -0.4686     0.2223  -2.108   0.0369 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4087 on 136 degrees of freedom
Multiple R-squared:  0.03163,    Adjusted R-squared:  0.02451
F-statistic: 4.443 on 1 and 136 DF,  p-value: 0.03688
```

Results:

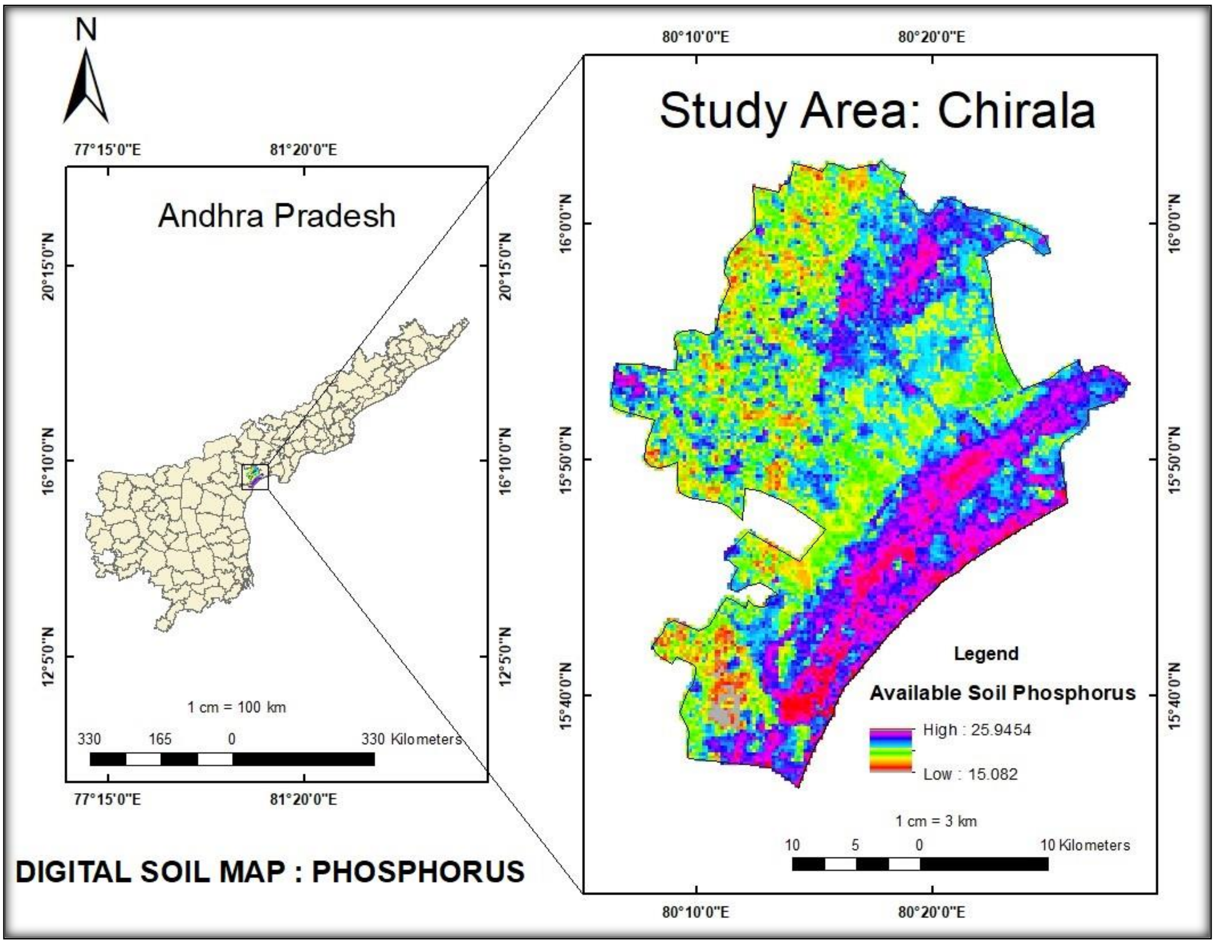
Using the regression model for the two parameters (P and Zn) the prediction of the values for these dependent variables is carried out and the nutrient distribution maps are generated.

Digital Soil Map:

A digital soil map is created using the values for the two parameters and is shown below,

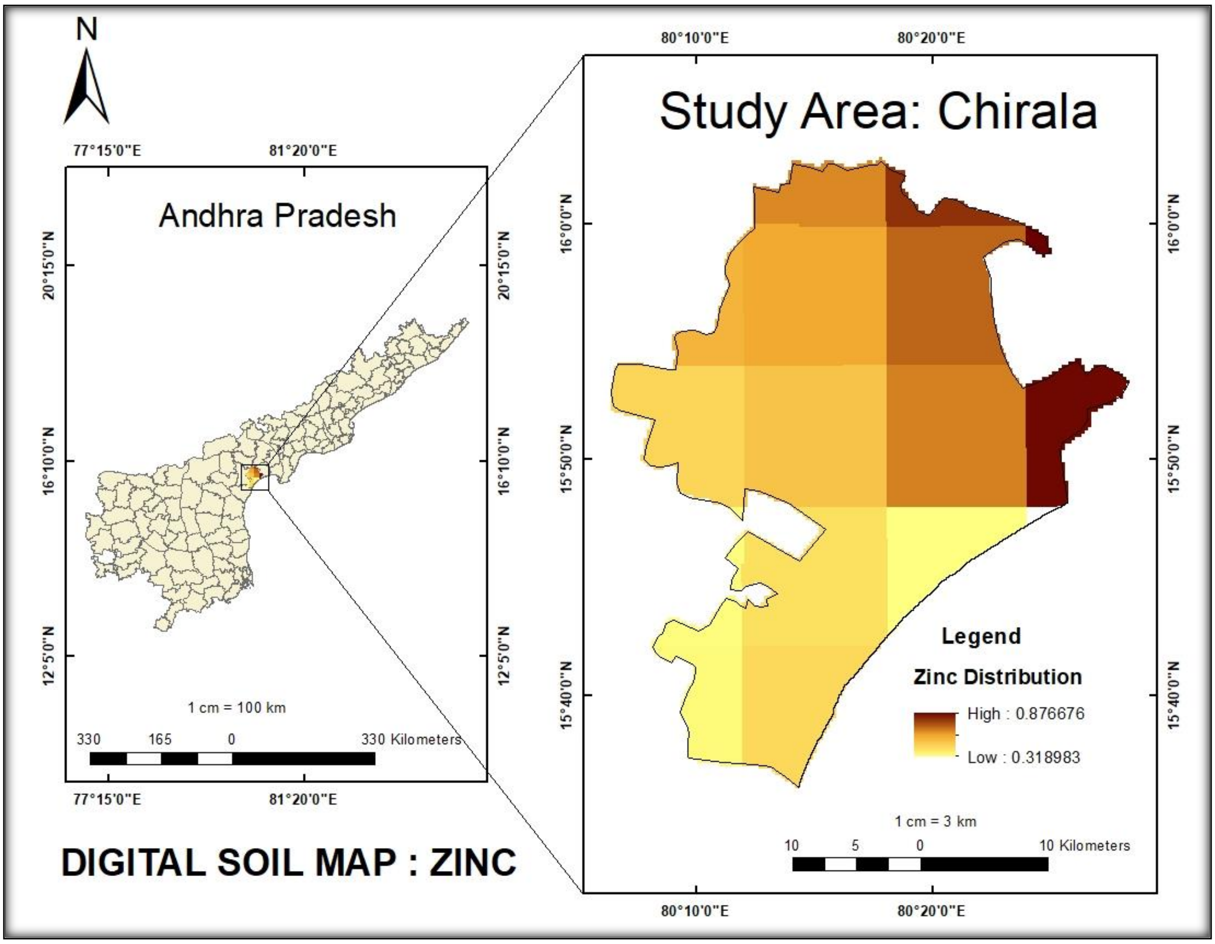
Soil map showing the distribution of Phosphorous.

Using the soil data of the nutrient phosphorous (2017-2019) in the ROI, the values are predicted for the year 2020. The below map shows the soil nutrient distribution map for phosphorous for the area Chirala.



Soil map showing the distribution of Zinc.

Using the soil data for the zinc nutrient in the ROI for the years (2017-2019), the values are predicted for the year 2020. The below map shows the soil nutrient distribution for zinc for the area Chirala.



The above categorization of the soil properties are based on the (INDIAN COUNCIL OF AGRICULTURAL RESEARCH) ICAR standards.

REFERENCE

[1] Digital soil mapping: A brief history and some lessons - ScienceDirect

[2] https://www.sciencedirect.com/science/article/pii/S0016706115300276

[3] AGRICULTURE | Prakasam District, Government of Andhra Pradesh | India

[4]Zinc for crop production | UMN Extension

[5] Why phosphorous is important (nsw.gov.au)