Soil Nutrient Estimation By Data Analysis to improve crop productivity of Indian Farmers

[1]N.Ram Kumar, [2]R.Vinay, [2]P.V.K.Gowtham, [2]M.V.K.Kalyan, [2]P.Anirudh
[1]Assistant Professor, [2]Student, [2]Student, [2]Student, [2]Student
[1]Electronics and Communication Engineering
[1]Anil Neerukonda Institute of Technology and Sciences , Visakhapatnam, India

Abstract : Crop productivity is a major issue which has been effecting the revenue of any agriculture sector. The method suggested mainly focuses on prescribing a solution for the productivity issues in crop which arise due to soil nutrient composition. One reason for improper productivity is unavailability of data with the farmer regarding the current status of soil and proper recommendations to improve the yield. In a hugely populated country like India it becomes a mere impossibility to test a piece of land before every season. So our mechanism is to device a mathematical tool which would give an estimate of the soil status before every crop season. The tool utilizes data available from previous soil tests and a small chemical test. The core mechanism used is solution of parabolic differential equation using explicit scheme. The mechanism prescribed would give an estimate of soil status to each and every farmer and recommendations for better yield thus improving productivity. As a part of mechanism, focus is put on error reduction techniques so the obtained estimate is as accurate as possible and equivalent to a full-fledged chemical test from laboratory.

Key Words- Mathematical, Productivity, Chemical Test, Differential equations, Explicit Scheme, Error reduction.

I. INTRODUCTION

India is an agriculture based country and nearly 60% of the population are involved in agriculture as their major occupation. The revenue from the agriculture sector is not in comparison to the population involved in it. One major reason for this revenue problems is improper productivity from crops. There are a number of factors resulting in productivity issues and one of the problems to be addressed is improper soil testing. Soil status plays a major role in crop and fertilizer selection. If this is not done properly, a drastic productivity drop can be observed. The main theme behind the suggested technique is to develop a mechanism which can effectively address this issue. It has been observed that as there are a large number of villages in comparison to the soil testing laboratories, it is not possible for them to provide a soil status report to every farmer before every crop season. So the paper discusses on providing a tool to resolve the problem. The mechanism suggested would act as bridge between the labs and the farmer. The technique for its functionality would utilize all the available data of any previous seasons with the labs and, for a current estimate in the user end that is the farmer end we use a rapid chemical test to bring both these results on to one single platform. These data entries act as inputs for computation mechanism suggested and finally arrive at an estimate of the soil nutrient values.

II. IMPLEMENTATION

A. Overview

Mathematics has been an area where different techniques predefined in it find new ways to be applied in different fields and here the technique extends the use of a mathematical technique in the agriculture domain. The mechanism as stated above is parabolic differential equation using explicit scheme. Differential equations one of the core areas in mathematics basically deals with finding solutions of small responses thinking in a broader perspective differential equations is predicting behaviour from a predefined set of rules . Thus it can be understood any prediction models which can be well defined can use the technique of differential equations to estimate long term responses. Differential equations are basically designed to deal with continuous curves and functions. Computationally the continuous function analysis is not possible. The computational equivalent of the differential equation topic in mathematics is the numerical methods. The discretised format of solving a differential equation can be observed in numerical methods. As our model tends to compute data we are utilising the numerical methodologies to solve discretised differential equation.

B. Differential equation in our technique

The model suggested tends to use the numerical solution of a parabolic differential equation. This mechanism is finalised by considering various factors. In any crop season, the farmer follows two processes. Firstly, the soil is fertilised which increases the nutrient level in soil and later the fertilization is followed by the process of plantation. Plantation consumes the nutrients of soil. Thus a decrease in nutrient level is be observed. If we can observe, there is a rise followed by fall which in mathematics represents a parabolic behaviour. So the possible response available from the crop is having a parabolic in behaviour.
Once the process is finalised, the next step is the constraints required to obtain a well-defined solution.

Parabolic differential equation has constraints of initial condition, boundary condition, positive slope and a negative slope. The various physical parameters available in the practical scenario are to be mapped to the constraints to obtain a solution by using the technique suggested.

III. MECHANISM OF IMPLEMENTATION

The mechanism is performed in three stages:
1. Solution for differential equation
2. Error estimation analysis
3. Mapping to ranges in chemical test

1.Solution for differential equation

The constraints are initial condition, boundary condition, increasing slope and decreasing slope. The mapping of these constrains in our scenario is made as per requirement. Every estimation forms a point of previous available data. This acts as the initial condition. Every soil has a critical value below which the nutrient level cannot fall. This acts as the boundary condition. The increasing slope is caused due increase in nutrient level which is caused due to the process of fertilization. The fertilization is done in random by the farmer wherein two possibilities arise. The farmer might have crossed the required level of nutrients or failed to meet the required nutrient level. Let us name these two cases as over fertilization and under fertilization respectively. The increasing slope is further subdivided into two variables comprising of the under fertilization range and the over fertilization range. Both these slopes are obtained by using a simple weighted sum algorithm on the data available with laboratories. The data of a particular village is considered then the data is passed through the weighted sum algorithm which would add a weight to the range in which a particular entry falls (if 6-15 is overall range with 6-10 is under fertilization range 10-15 is over fertilization range if it is divided into blocks of 0.25 interval if a random value from data set is 7.6 then the column 7.5-7.75 would have its weight raised by one). After the entire data set passes through this mechanism, we have respective weights of each sub block. The mean of mostly mapped sub block in the under fertilization range is used for one slope and the mostly mapped sub block in the over fertilization range is used for the second slope. Both these slopes collectively form the positive slope estimate. The positive slope variable is not only an estimate of type of fertilization but also provides an estimate of the geographical soil effect as for each analysis, an entire village data is considered and the trends would depict the geographical behaviour also. The final constraint required is the negative slope. It is a parameter depending on plantation or nature of the plant. Each and every plant has a different pattern of consumption of different nutrients. If these consumption level graphs for different plants are discretised then we would obtain the negative slope patterns for different plants. These four physical constraints if applied to the parabolic solution, it would then produce the required data of the current soil status.

One important fact to be considered here is that, when we are mapping to the highest weighed sub blocks in positive slope determination, there are other blocks with entries mapped which are being neglected. The effect of these entries would result in an error for the estimated answer. Thus, it is important to calculate the error of each estimated result so that we can arrive at a much accurate result.

2. Error estimation analysis

As stated above, error estimation is required for an accurate analysis. The error estimation method that is being implemented as a part of the technique is the Gaussian error approximation graph. The Gaussian curve basically requires a mean and a standard deviation to be plotted. We have the most mapped sub block available which acts as the mean of the Gaussian estimate. Every plant has a range of consumption under which there is no significant effect on their growth. This would give us the range for standard deviation for the estimate. The Gaussians are plotted using the mean and standard deviation values. The area behind the standard deviation would form the error of the approximation made. Thus the error value for each mapped value is obtained.

3. Mapping to ranges in chemical test

The calculations made so far are approximations from the available data base. But for any estimation to be successful a hint of current status is required which is here performed by a rapid chemical test. The rapid chemical test comprises of capsules to measure the Nitrogen, Phosphorous and Potassium levels in soil. When a mix of soil is added with a capsule, it would readily turn into a colour and based on the colour shade, subdivisions exists like trace, low, high, very high levels of nutrients in the soil. Now consider the numerical estimate to be performed for two crop seasons. The combinations obtained would be four (under fertilization followed by under fertilization, under fertilization followed by over fertilization, over fertilisation followed by under fertilisation and finally over fertilisation followed by over fertilisation). Each of this would result in a different range value. Out of all the possibilities, the most feasible answer is to be suggested. This is done by mapping the results of all the combinations to the different levels of the chemical test. If a one to one mapping is framed between the results then the answer in the range of chemical test range is segregated. If a many to one mapping occurs in result, there are more than one possible values in the chemical test range then the segregation is done based on the error estimation analysis. The result with higher error is removed and the value occurring with lower error is given as output. Thus from the above three steps we are able to obtain an estimate of current soil status using Numerical methods and error is minimised using Gaussian error estimate.

IV. RESULTS

The histogram represents the data set considered for analysis plotted as an histogram which comprises of the number of times each and every sub block is been mapped.

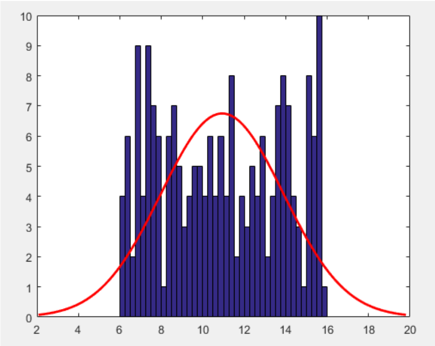


Fig: Histogram of the dataset

The table indicates the numerical computation table. The starting value is the available data value which has been evaluated as we move in the table. The entry present in (7, 5) location is the next estimate for the value which is again used as input for next table as indicated and the further estimation is made. (In the above example, an estimation from 6 to 6.6685 is made considering two seasons of under fertilisation with the required value as 10). The table basically is a step wise representation of how a parabolic differential equation is solved.

7.3750

6.0000	6.4583	6.9167	7.3750	6.5750
6.0000	6.4583	6.9167	6.7458	6.8750
6.0000	6.4583	6.6021	6.8958	6.5604
6.0000	6.3010	6.6771	6.5813	6.7104
6.0000	6.3385	6.4411	6.6937	6.4745
6.0000	6.2206	6.5161	6.4578	6.5682
6.0000	6.2581	6.3392	6.5422	6.3913

7.6250

6.3913	6.8025	7.2138	7.6250	6.8250
6.3913	6.8025	7.2138	7.0194	7.1250
6.3913	6.8025	6.9109	7.1694	6.8222
6.3913	6.6511	6.9859	6.8666	6.9722
6.3913	6.6886	6.7588	6.9791	6.7451
6.3913	6.5751	6.8338	6.7520	6.8388
6.3913	6.6126	6.6635	6.8363	6.6685

Table: The iterations involved while numerical

The graph below is an error estimation graph. It comprises the mean followed by a deviation range and the marked area in the graph is the unfavourable area. The error caused by the slope determination is stated above. As the process of computation occurs, each stage answer with the error from all the previous levels are displayed. Finally the mapping and sorting algorithm sorts the best feasible answer from all the available estimations made.

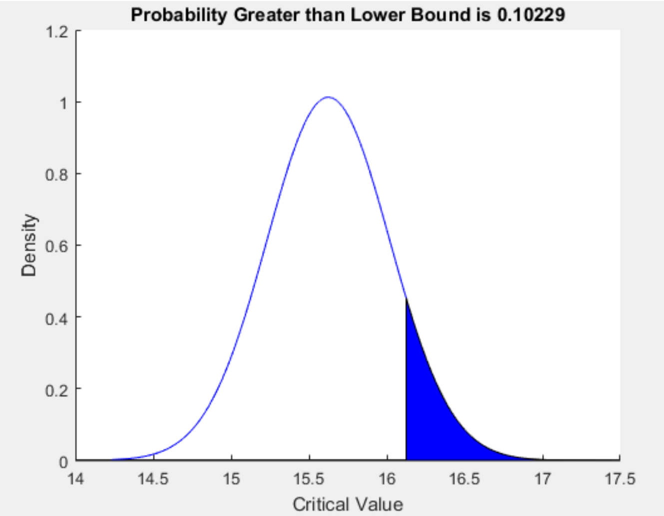


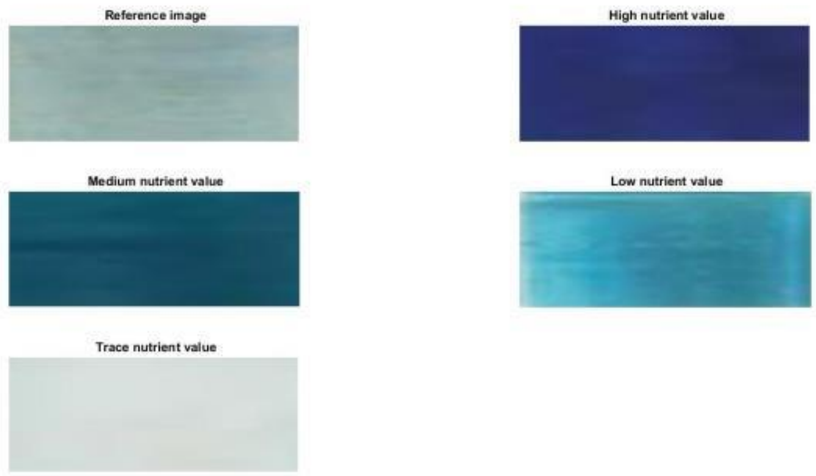
Fig: Error Estimation

The estimated answers with their error will be as follows with initial value 6 the under fertilisation followed by under fertilisation is to produce a value 6.6685 with an error 0.0022 under fertilisation followed by over fertilisation is to produce a value 12.5214 with an error 0.0057 over fertilisation followed by under fertilisation followed by under fertilisation is to produce an error of 0.1510 and finally over fertilisation followed by over fertilisation is to produce a value 14.6121 with an error of 0.3331 for the above example .

```
 6.6685

 0.0022

12.5214

 0.0057

 9.3647

 0.1510

14.6121

 0.3331
```

Image: The results of the 4 possible cases with their error probability

Chemical Test



Chemical tests are performed and the images are processed using MATLAB to segregate an answer from all possible outputs obtained above.
The test image matches with Low nutrient level. Thus from the answers above, *9.3647* will be the nutrient level of the soil.

V. *METHOD*

a) ***Solution for parabolic equation****::* The parabolic equation discussed above is being implemented using the below molecular scheme.

	1		j+1
r	(1-2r)	r	j
i-1	i	i+1	

The formula is obtained from the above molecular scheme for explicit method for solving the parabolic difference equation:

$$u_{i,j+1}=ru_{i-1,j}+(1-2r)u_{i,j}+ru_{i+1,j}$$

The explicit scheme is chosen mainly due to its compatibility and stability. The condition for convergence is the value of ‘r’ must be less than or equal to 0.5.

CONCLUSION

The above process demonstrates a mechanism that can estimate the current soil status from the data available and a chemical test. The method also includes an error estimation analysis which would improve the accuracy of results to a large extent. The suggested method acts as bridge between the labs and farmers whenever the labs are unable to perform a complete chemical test of soil this estimation mechanism can be used to give an estimate at the farmer end so that the farmer gets a complete idea of what he has in his hand. The functionality of the following method can be effectively used by creating a webpage which links the data available from labs to the rapid chemical test result from farmer. Village level centres may be established to update the results of the rapid chemical test and the created webpage after receiving all the inputs would perform all the required computations as per the mathematical model suggested and produce an answer which would be present soil status report. This would help the farmer take right choice of fertiliser and crop. This results in precision agriculture being implanted. Finally there would be an effect on the yield of crops resulting in increased productivity. Increased productivity would in turn increase the revenue of the agriculture sector finally resulting in improvement in revenue of country thereby achieving sustainable development of the country.

REFERENCES

[1] Higher Engineering Mathematics by B.S.Grewal 43rd edition

[2] Numerical Methods-Finite Difference Approach
https://onlinecourses.nptel.ac.in/noc18_ma03/

[3] Probability, Random Variables and Random Signal Principles by Peyyon Z Pebbels, Jr 4th edition

[4] The numerical solution of parabolic and elliptic differential equations
DW Peaceman, HH Rachford Journal of the Society for Industrial & Applied Mathematics Vol 3, No 1

[5] http://www.agrinex.in/soil_doct_plus.html

[6] http://www.jocpr.com/articles/nutrient-analysis-of-soil-samples-from-various-places.pdf

[7] http://makeyourownneuralnetwork.blogspot.com/2015/01