Machine Learning for Soil and Crop Management Professor Somsubhra Chakraborty Agricultural and Food Engineering Department Indian Institute of Technology Kharagpur Lecture 25 Diffuse Reflectance Spectroscopy: Basics and Applications for Crop and Soil (Contd.)

(Refer Slide Time: 0:18)



Welcome friends to this twenty fifth lecture of NPTEL online certification course of Machine Learning for Soil and Crop Management. And we are at week 5 and this is the fifth lecture of week 5. And in this week, we are talking about the Diffuse Reflectance Spectroscopy: Basics and Application for Crop and Soil. And in our previous 4 lectures, we have seen the basics of diffuse reflectance spectroscopy specifically, the visible to near infrared diffuse reflectance spectroscopy, and then we have seen the basics of some the working principle of spectroradiometer, some specification of the spectroradiometer.

Then we have also seen the spectral resolution, sampling interval, spectral pre-processing methods and we have also seen the important spectral features, some of the important literatures in the domain of DRS application for soil. So, guys, today, we will start from there, and then we will be discussing some more application on soil and also we will be discussing some application on crops.

(Refer Slide Time: 1:40)



So, these are the 2 concepts which we are going to cover VisNIR DRS soil application and VisNIR DRS crop application in this lecture and these are some of the keywords which we are going to discuss today. And then soil arsenic pool, then orthogonal signal correction or OSC, then external parameter orthogonalization or EPO, RSI, NDSI are 2 index which are used for predicting some of the crop parameters via the spectral modeling we will be also seeing.



So, let us see some example of let us first see some example of DRS based, sorry VisNIR DRS soil application. So, in our previous lecture, we have seen some of the literature where DRS was used for predicting several soil fertility and also soil physiochemical properties. DRS has also been used for predicting the soil heavy metals, metalloids as well as other contaminants. So, let us see some example.

So, this is one example where DRS was utilized for soil arsenic prediction. So, this study actually was conducted in a landfill nearby Kolkata city called Dhapa. And several hundreds, couple of hundreds of soil samples were collected from these sites and after collecting the samples, we have scan them using the diffuse reflectance spectroradiometer. So, once we have collected the soil samples, we scan them using the VisNIR spectroradiometer. And then we estimated the total soil arsenic using the standard method which we have used as a response soil arsenic values and just like the spectral transfer model, we have used the VisNIR DRS there is reflectance spectra as the predictor.

But before using in the model, we have done first derivative reflectance spectra, then second derivative reflectance spectra, then orthogonal signal correction these 3 are the spectral preprocessing and after the spectral preprocessing, we have used the elastic net model, Random Forests model and penalize spline regression model. And then, we have evaluated the model based on the validation set and we have selected the best performing model and then we have used the model predicted arsenic values. And finally, after we predicted the arsenic values using the best prediction model, we have interpolated those values using ordinary kriging and indicator kriging, which are geo statistical approaches. So, these geo statistical approaches are useful to predict the parameter in some in those locations where from un-sampled location. So, these interpolations were done by ordinary kriging and indicator kriging.

(Refer Slide Time: 5:26)



So, we have seen this is the distribution of the major arsenic concentration and when we have used the elastic net model, we have seen we are getting the best prediction accuracy with an R squared values of 0.97 and RPD values of 6.32 which is very high. So, using the elastic net model, we have predicted the arsenic with high accuracy. So, that shows that these arsenic can be predicted by in the soil using diffuse reflectance spectra.



So, these are the original reflectance spectra of the soil for those samples and these is the elastic net coefficients and these red dots are showing the significant wavelengths which have been selected by the elastic net model. Now, using this elastic net regression model, when we predicted the arsenic, we have also interpolated those values using the indicator kriging as well as the ordinary kriging. So, these are the measured and predicted arsenic concentration using ordinary kriging and these are the measured and predicted arsenic soil arsenic content using the indicator kriging.

Indicator kriging generally gives you maps the spectral variability based on the probability of any value which is exceeding a threshold value. So, in our case, we have kept the threshold values at 5 ppm or 5 milligram per kg and we have tried to identify the areas which are having 0 to 25 percent probability, 25 to 50 percent probability, 50 to 75 percent probability and 75 to 100 percent probability and these greens areas are water bodies. So, you can see in both the cases in case of the measured values as well as the predicted values we have seen, we have identified the similar hotspots.

So, these regions are the pollution hotspots. So, the identification of this pollution hotspot by this kriging interpolation can help you to identify the areas where you should give special care and so, similarly, in case of indicator kriging also, these 2 areas have been come these 2 area came under this 75 to 100 percent probability showing the potential hotspot zone.

So, that implies that our DRS method was able to identify the DRS method was able to identify or predict the soil arsenic content not only they have they were able to predict the

soil arsenic content contamination, but also they were able to produce the almost accurate special variability map or using the geo statistical interpolation like ordinary kriging and indicator kriging.

(Refer Slide Time: 9:13)

	Analysis of soil solid As phases has	commonly used seque	ntial extraction: Time con	suming,								
	destructive, and costly											
	VisNIR DRS spectral data for rapidly (predicting total As and	five different solid As phas	es (Mø								
	visinin ons spectral data for rapidly predicting total As and five different solid As phases (wg,											
	PO4, Ox, HCI and org pools)											
Table 1	hid astronion secondus for it-basis cald abuse The lot abus (Dm.)	une entructuel convertable (conceiled from 3 h	none at al. 1007 and Know at al. 2009).									
Table 1 Sequer Step	n Inial extraction procedure for As-bearing solid phases. The last phase (Og) Target phase	was extracted separately (compiled from Ah Extractant	nam et al., 1997 and Keon et al., 2001). Possible mechanism	_								
Table 1 Sequer Step Mg	til till extraction procedure for As-bearing sold phases. The last phase (Og) Target phase Ionically bound As	was extracted separately (compiled from Ah Extractant 1 M MgCl ₂ pH 8	nann et al., 1997 and Keon et al., 2001). Possible mechanism Anion exchanism (CI for As; possible Mg-As	=								
Table 1 Sequer Step Mg PO ₄	l al attraction procedure for As-bearing sold phases. The last phase (Og.) Target phase Inscially bound As Strongly adouthed As	was extracted separately (compiled from Ah Extractant T M MgGs pH8 T M NaH ₂ PO ₄₀ pH5	rann et al., 1997 and Keon et al., 2001). Possible mechanism Anion exchange of Cl for As; possible Mg-As complex formed Anion exchange of DQ, for KeO2, and AsO3	-								
Table 1 Sequer Step Mg PO ₄ HCI	I detraction procedure for A-bearing solid phases. The last phase (Org.) Target phase Institute of the operation of the opera	was extracted separately (compiled from Ah Extractant 1 M MgCl ₂ pH 8 1 M NaH ₂ FO ₄ pH 5 1 N HAT	ann et al., 1997 and Keon et al., 2001). Prosible mechanism Anion exchange of CI for As; possible MgJ6 complex formed Anion exchange of CO ₀ , for ACO ₀ and AGO ₃ Proton dissolution; Fe-CI complexation									
Table 1 Sequer Step Mg PO ₄ HCI	ful detraction procedure for Arboaring solid phases. The last phase [Og] Target phase Insticulty bound Arb Strangty advocable Arb Aroproprintized with andronaess. Me middes, and ery anorphous fer supplyamises Aroumentizated with anorphous for constraintsings.	was extracted separately (compiled from Ah Extractant 1 M MgCb ₂ pH 8 1 M NatJb ₂ Op H 5 1 N HGL 0 2 M ammonium notately intalic add	nam et al., 1997 and Keon et al., 2001). Prosible mechanism Anion exchange of Of Ior Ac; possible Mg-As complex formed Anion exchange of Oo ₂ for KoO ₂ and AGO ₂ Proton disslution; FeGL complexation Literack ensurer of elsowhriton	-								
Table 1 Sequer Step Mg PO ₄ HCI Ox	nd etración procedure for Asbaning sold phases. The last phase (Og) Target phase Innically bound As Strongly adambed As As compresiption with adomates, Mo noide, and ety anorphono. Fe anylphonicks As compresiptioned with amorphones Fe engly-thronides	was estracted separately (compiled from Ab Estractant 1 M MgCs pHS 1 M Mail-So pHS 1 N HCl 0 2 M ammonium noulare (snalic assi, pH 3.	ann et al. 1997 and Koon et al. 2001). Possible mechanism Anton exchange of Cl for Acy possible Mg-46 complex formed Anton exchange of Olg. for AcQ, and AcQ, Proton dissultation; Fe-Cl complexation Ligand-promoted dissolution									
Table 1 Sequer Mg PO ₄ HCI Ox Org	fal etraction procedure for Arboaring solid phases. The last phase (Dg) Target phase Instically board Arb. Stronging also boards Arb. Ar converginated with arboarates, Mo moldes, and ery samphonade Arc converginated with armophone Ferophysionides Are converginated with armophone Ferophysionides Are somewhat and armophone Ferophysionides	was extracted separately (compiled from Ah Extractant 1 M MgCb ₂ pH 8 1 M MdFb ₂ pH 8 2 M attraction of the second seco	ann et al. 1997 auf Ken et al. 2001). Pouble mechanism Anion erchange of Gin Ariz pouble Mg-As compared formed Anion exchange of PO ₂ for AO ₂ and AO ₂ Protein dissolution: Fe C complexistion Ligand-promoted dissolution -									

Another in the subsequent study, we have tried to predict the soil arsenic pool. Now, that analysis of soil solid arsenic phase has commonly used different types of sequential extraction which are time consuming which are invasive or destructive and also very much costly. So, in the soil, arsenic remains in different pools. So, for example one of the form is called we call them Mg pool, which is basically the inorganically bound arsenic. And then there is a pure PO4 pool or phosphate pool which is strongly absorbed arsenic.

And then there is HCl pool which is co-arsenic which is co-precipitated with the carbonates, manganese oxide and very amorphous iron oxy hydroxides and OX is the arsenic coprecipitated with the amorphous iron oxy hydroxides and organic is arsenic associated with the organic method and sulfides. So, you can see in the soil, arsenic can be associated with different phase and for proper identification of this phase or proper prediction of this phase is very much important if we want to map that dynamics of soil arsenic.

So, can we map it using diffuse reflectance spectroscopy that was the hypothesis this was the research question. So, our hypothesis was yes, we can do using the diffuse reflectance spectroscopy and appropriate spectral pre-processing. So, generally these individual target phase are extracted generally in conventionally they are extracted by using different types of extracted and it is a very cumbersome process. So, can we use VisNIR DRS as an alternative

to prey dig these individual pools of arsenic in the soil along with the total arsenic? So, that was our objective or that was our research question.

(Refer Slide Time: 11:30)

Parameter	Spectral preprocessing	Calibration R ²	Calibration RMSE (mg·kg ⁻¹)	Validation R*	Validation RMSE (mg·kg ⁻¹)	RPD ²	Bias	RPIQ	
Ast /	10	0.95	0.35	0.88	0.57	2,85	0.01	3.54	
	2D	0.91	0.47	0.86	0.59	2.73	0.01	3.37	1964 C
1	OSC -	0.87	0.60	0.78	0.75	2.15	-0.05	2.66	
Org	1D	0.97	0.22	0.93	0.35	3.74	-0.004	4.81	
	2D	0.94	0.33	0.89	0.43	3.11	-0.03	4.00	
1	OSC	0.88	0.48	0.76	0.65	2.05	-0.05	2.64	
Mg	1D	0.67	0.01	0.25	0.01	1,16	0.002	1.77	
	2D	0.59	0.01	0.34	0.01	1.24	-0.0007	1.89	
1	OSC	0.60	0.01	0.27	0.01	1.18	0.003	1.80	
PO4	1D	0.85	0.36	0.72	0.50	1.91	-0.003	3.39	
	2D	0.83	0.41	0.71	0.51	1.88	-0.025	3.33	
1	OSC	0.82	0.42	0.64	0.56	1.69	-0.03	2.99	
на	1D	0.88	0.03	0.587	0.05	1.56	0.01	2.32	1000
	2D	0.83	0.04	0.59	0.05	1.57	0.01	2.34	
1	OSC	0.79	0.04	0.52	0.06	1.45	0.01	2.15	1 63
Ox 🖌	1D	0.62	0.50	0.46	0.60	1.38	0.01	2.09	1 may
	2D	0.58	0.53	0.38	0.65	1.28	-0.02	1.95	Charles .
	OSC	0.57	0.53	0.44	0.62	1.34	0.03	2.04	

So, what we did? We after we measured these individual phase individual target phase, then we have used the arsenic soil spectra to predict using different model different PSR model. So, you can see here there are at 3 different spectral pre-processing we have used, first derivative, second derivative and orthogonal signal correction and this is AST means total arsenic, then organic pool, Mg pool, PO4 pool, HCL pool and OX, OX pool.

This shows the calibration R square, calibration RMSE and validation and you can see comparing all of these that in case of first derivative, second derivative total arsenic was predicted very high, organic also very good, Mg pool was not that highly predicted. In case of PO4 pool it was satisfactorily predicted, in case of HCl pool it was moderately predicted. So, you can see the prediction accuracy changes from one pool to another pool, but most of the pools are satisfactory predicted except for a one or two.

So, that shows that a combination of PLSR and a proper spectral processing can identify these individual phases of arsenic in the soil, which may help in future identification of the soil arsenic dynamics. So, that was also published in another journal called the Geoderma.

(Refer Slide Time: 13:15)



So, VisNIR DRS for these are the prediction plots you can see our scalar values for total arsenic 0.88, 0.92, 0.58, 0.70, 0.83, 0.54. So, these are the prediction plots for different arsenic pools in the soil using diffuse reflectance spectroscopy.

(Refer Slide Time: 13:39)



Now, considering the complexity of these pools, these values are still very much encouraging because, you remember that this soil arsenic and this complex and their phases are at so much heterogeneous and it is very much if we can get a moderate prediction also for some of them, that would also help us that is also greatly helped us for identifying their dynamics in the soil.

These are the spectral the coefficients, PLSR coefficients and these PLSR coefficients have identified these red dots, which are the significant PLSR coefficients which may use very useful in future research if they want to pinpoint, if they want to selectively use these wavelengths for a subsequent machine learning application.

(Refer Slide Time: 14:43)



I have already shown you this application of DRS for petroleum contaminated soil so I am not going to discuss it again. But just briefly you can see that this petroleum contaminated soil using the diffuse reflectance spectroscopy and also you can see the what we have used original reflectance first derivative and second derivative in different condition like field moist intact, air dried intact and air dried ground. So, we have got in terms of validation at (()) (15:19), we have got satisfactory R squared values for field moist intact when you considered the validation R squared and also the potential RPDs were also encouraging.

So, the RPD values were 1.64, 1.70. So, if we can focus, if you can see that first derivative of the reflectance spectra using the field moist intact gave the best reading when we consider the validation R squared among all the models and also the RMSE values and also the RPD. So, using these we can consider that DRS is a useful method for identifying the petroleum contamination in soil.

(Refer Slide Time: 16:04)



Another very important application for soil is, you know that soil contain moisture and moisture absorb this infrared radiation. So, one of the major obstacle for in situ application for diffused reflectance spectroscopic based soil property modelling is the presence of moisture. So, to remove the impact of moisture, scientists have proposed a methodology that is called EPO or external parameter orthogonalization. It was first proposed by Roger et al in 2003.

And this EPO algorithm is based on the proposition that field moist spectral matrix if we consider this an X can be expressed as an amalgamation or mixture of an useful section which is denoted by XP that describe a target soil properties suppose it is clay and extraneous section using the influence of moisture which is XQ and also spectral noise. So, you can see that the field moist spectral matrix can be decomposed into these 3 components.

This is spectral noise and this is the external section including the influence of moisture XQ and this is the XP which is target soil properties up in our case it was clay. So, the useful part this is only the useful part can be computed by the difference matrix, what is the difference matrix D, which is the difference between the spectra of the soil with moisture and spectra of the dry soil. So, if we have taken the spectra from moist soil and the same soil which is dried, then we can differentiate the spectra that will give you the difference matrix D.

Note that the Q can be calculated from the singular value decomposition which is known as SVD of D and subsequently can be used for deriving the projection matrix P from this equation, where I denotes the identity matrix. So, specifically this SVD helps in removing the

space of the moisture ultimately generating an orthogonal space that remains unaffected by the moisture effect. So, in other way, in other words, basically this EPO tries to orthogonally project the useful spectral section from that of the moisture from that section which is influenced by moisture or that is XQ.

So, when then the 2 variables are orthogonal projected from each other, then they do not have any interaction among this. So, in this way by orthogonally projecting the spectral space using this mathematical procedure of EPO, we can remove the impact of moisture. Notably, the number of EPO components can be derived from the SVD of D, alternatively, Q can be calculated by the principal component analysis of D transpose D, this is D transpose D. So, it should be read as D transpose D.

So, if you do the principal component analysis you can also calculate the Q from this principal component analysis. So, the idea is EPO is useful to orthogonally project the useful part from the non-useful part that is moisture and in that way we can remove the impact of moisture for in situ prediction of soil properties without the influence of moisture.



(Refer Slide Time: 20:07)

So, we have used this technology for several scientists have used this technology and we have also used this technology. And we have seen that our result shows that using EPO method, it was possible to reduce the error introduced by the weight or in situ measurement. So, if we can utilize this EPO, we can reduce the error which we can get when we scan the sample in situ of course, because in situ samples are having moist moisture and this moisture can produce some errors. And that error can be reduced if we go with this external parameter orthogonalization.

(Refer Slide Time: 20:52)



Scientists have also used a penetrometer mounted VisNIR probe for collecting the VisNIR spectra you can see here Jason Ackerson et al in 2017 they have used thiss penetrometer mounted VisNIR probe this is the original picture and the exterior, this is an exterior picture and you can see this is an interior picture. So, this here is a lamp, here is a mirror and ultimately this is a fiber optic cable. So, as this penetrometer goes inside the soil, the fiber optic cable is getting the soil scan, so, for from in situ soil scan.

So, for each sampling location visitor spectra were collected in situ using the penetrometer mounted probe and from air dried and ground soil course. So, and then followed by an EPO was applied to in situ VisNIR spectra to remove the effects of the soil moisture and intactness of the spectra. And after application of the EPO, PLSR model calibrated using a spectral library containing only spectra from these air dried and ground soil were used to predict the clay content from in situ spectra.

Ultimately, they have seen that prior to application of EPO, PLS was unable to accurately predict clay content from the in situ spectra. However, after application of the EPO, PLS performance or in situ spectra was quite good with an average RMSE and R square across all sampling areas 88 grams per kg and then 0.0, asked values of 0.76. So, it shows that application of EPO in this VisNIR penetrometer mounted VisNIR probe also gave very good

R square and remove the inherent error which you can get from scanning the soil sample in situ, moist samples in situ.

(Refer Slide Time: 22:56)

difference spectral indices (NDSI) sensitive to sucrose, reducing sugar and total sugar content were identified which were subsequently calibrated and validated. The best performing multivariate models for sucrose, reducing sugars and total sugars were found to be, MARS, ANN and MARS, respectively with respect to RPD values of 2.08, 2.44, and 1.9.	Mentifying Suitable Bands, Indiers and Spectral models opposed index (ISSI, NDSI) Development of Prediction Models and Evaluation System Development of Prediction Models and Evaluation Spectral Pleasetyping for Water Deficit Stress		
---	--	--	--

So, let us see some examples for crop you can see here the use of DRS and chemometrics for water deficit in rice this research was done in a Das et al in (2018) was published by Das et al in 2018. So, what they have done, so, they did a field experiment where we the 16 genotypes and 3 replicates and then they have made the spectral measurement using FieldSpec portable spectrophotometer. And simultaneously they have measured the sucrose reducing sugar and total sugar and then they have identified the suitable bands, indices and spectral models.

They have proposed to novel indices, RSI which is the ratio of spectral index and normalized different spectral indices, NDSI and then they have used different models like artificial neural network, MARS, MLR, PLSR, RF and SVR and then they develop the prediction model and then evaluated the prediction model and then they have used the spectral phenotyping for water deficit stress.

So, ultimately, the best performing multivariate models for sucrose reducing sugars and total sugars were found to be MARS, multivariate adaptive regression spline, then artificial neural network and also MARS and also they produced the RPD values of 2.0, 2.44 and 1.9. So these RPD values are satisfactory. So, this shows the application of DRS and chemometrics can be useful for identification of the water deficit in rice.

So, they have developed these novel spectral indices in near infrared range, one is RSI, which is ratio spectral index and normalized difference spectral indices, NDSI, which is sensitive to sucrose, reducing sugars and total sugar content also identified which were subsequently calibrated and validated.

(Refer Slide Time: 25:08)



So, another application so, these are the spectral signatures you can see variation in spectral signatures in the rice leave at different waters deficit stress you can see these green one is controlled and the vegetative stress is this one and this is a flowering stress. So, you can see clearly see the difference in spectral patterns, when you go from the control to vegetative stress to flowering stress and they have predicted the sucrose and also different other parameters also using different models DRS. So, that shows that use of DRS and chemometrics can be useful for water deficit identification in rice.

(Refer Slide Time: 26:02)



They have also recently shown that use of DRS and chemometrics can be useful for salinity stress of rice also. So, this VisNIR DRS was used for leaf nutrient estimation in rice under salinity stress and novel spectral indices in VisNIR range were developed and validated and PLSR and PCA based machine learning models were evaluated and PLSR based models provided the superior performances for predicting the salinity stress in rice. So, that shows that different types of abiotic stress can be identified when we combine this DRS with the chemometric algorithms.

(Refer Slide Time: 26:46)



A recent research published by Brackx et al., they have shown that DRS in crop, diffuse reflectance spectroscopy can be utilized in crop for air quality monitoring. So, in urban areas, there is air quality is one of the major issues. So, to assess the potential of hyperspectral, this research were trying to assess the potential of hyperspectral tea leaf reflectance for monitoring the traffic related air pollution.

So, they have collected 29 Carpinus betulus sapling, they were exposed to the environment with either high or low traffic intensity and the local air quality was estimated by leaf saturation isothermal remanent magnetism and lift chlorophyll content, index specific leaf area which is denoted by SLA and water content were determined. And to gain insight in the link between the leaf reflectance and air quality, the correlation between this SIRM and several spectral features was determined.

And spectral reflectance for wavelength bands in the red and SWIR around the red edge were correlated with the SIRM with a Pearson correlation which you can see very high Pearson correlation coefficient 0.85 negative Pearson correlation coefficient negative correlation was there, but with the R squared value 0.72.

So, this is that shows that if we can use the diffuse reflectance spectroscopic based for capturing the reflectance from the leaf to identify the air quality. There is a, this is very novel application of DRS in crop.

(Refer Slide Time: 28:44)



One of our research in back in 2017, we have seen that if we can use the diffused reflectance spectroscopy for algae different parameters for predicting the different parameters in the algae also. So, for rapid assessment of algal biomass and pigment content also we have used diffused reflectance spectroscopy and PLSR model has produced very good results when we try to predict the chlorophyll A, chlorophyll B and then total carotenoids prediction for 2 different algae and we have got very good R squared. So, that shows the high, wide range of application of DRS both in soil and crop.

(Refer Slide Time: 29:30)



So, guys, these are the references, I hope that you have gathered some good knowledge, but of course, it is not possible to cover all these application, but I hope that I was able to pinpoint some important advances at important advancement in the field of DRS application for soil and crop. Of course, this will give you the impetus to go ahead and read more papers there are thousands and thousands of papers in the field of diffuse reflectance spectroscopy application in soil and crop.

So please go ahead and read more and more papers to gain more and more knowledge. And these things will be much clearer to you once you go ahead and read those papers. So, some of these references which I have utilized in this lecture are given here. But, of course, you are requested to go ahead and read some of the papers. And if you find any difficulties, let me know, I will be more than happy to answer your queries.

(Refer Slide Time: 30:32)



So, thank you guys. This makes the end of this week 5, and in the next week, we will start with a new topic. Thank you.