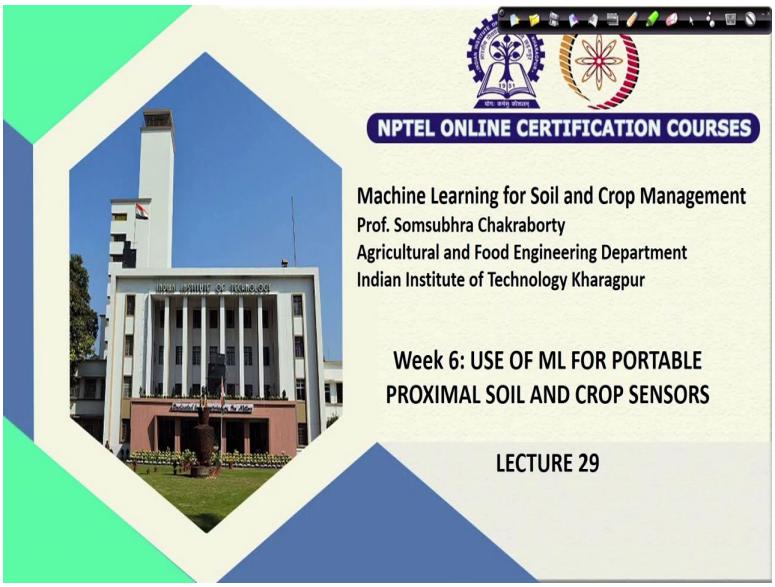
Machine Learning for Soil and Crop Management
Professor. Somsubhra Chakraborty
Agriculture and Food Engineering Department
Indian Institute of Technology, Kharagpur
Lecture 29
Use of ML for Portable Proximal Soil and Crop Sensors (Contd.)

(Refer Slide Time: 00:21)

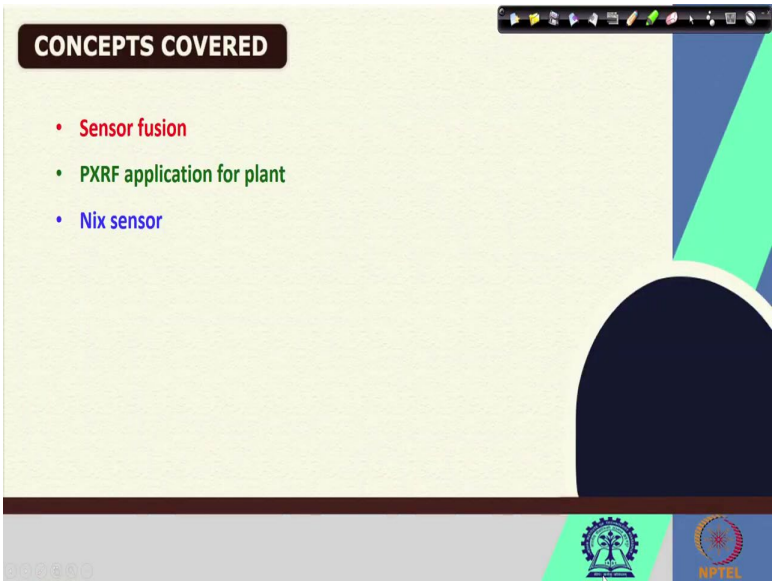


Welcome friends to this fourth lecture of week 6 of these NPTEL online certification course on Machine Learning for Soil and Crop Management. And in this week, we are talking about use of machine learning for portable soil and crop sensors. And in our first three lectures, we have discussed about the proximal soil sensors. And then we have discussed about site specific nutrient management smart soil sensing network.

And then we have talked about the portable XRF. We have seen the evolution of portable XRF from using the simple statistical model to machine learning model, their application in different aspects of soil, different physical chemical properties of the soil and how they are very expert or very much helpful for predicting different types of soil parameters we have seen.

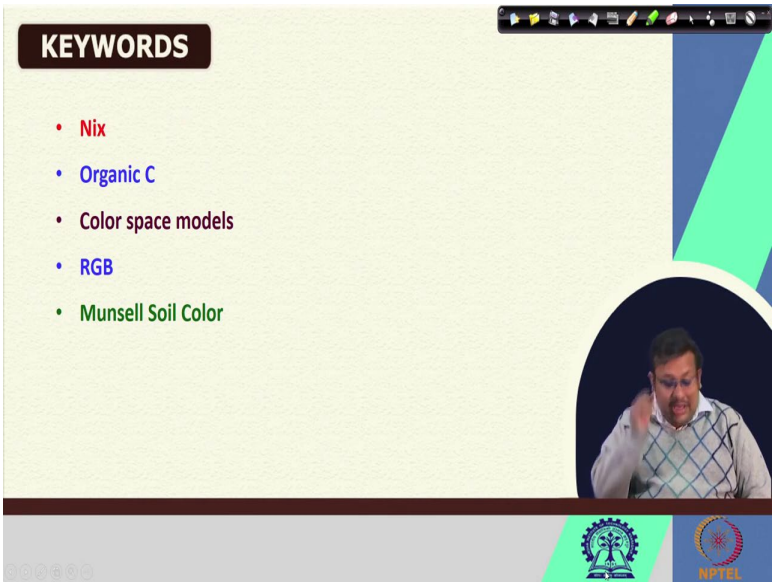
So, we will continue from there, we have started discussing about model fusion between the diffuse reflectance spectroscopy and PXRF. And today, we will go from there. And we will discuss some applications of model fusion.

(Refer Slide Time: 01:51)



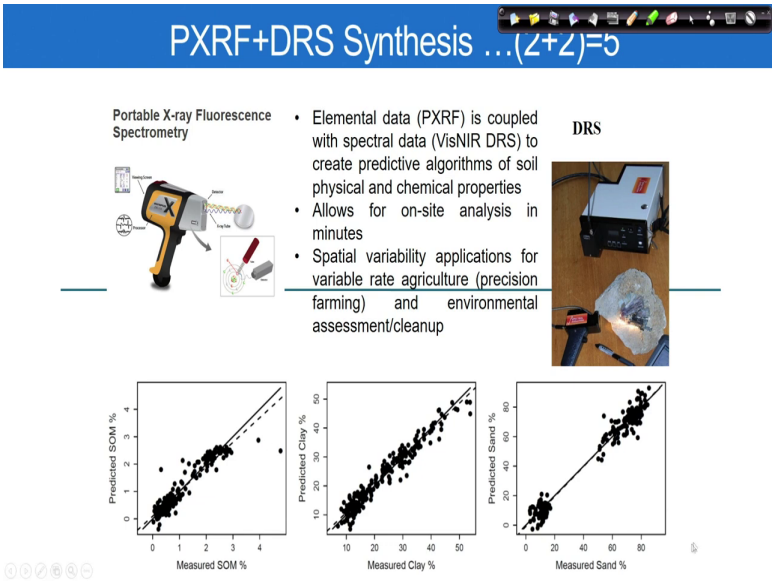
So, if you see the concepts, which are going to cover in this lecture are merely sensor fusion, then I will be showing you some application of PXRF for plant elemental content measurement. And then we will be starting discussing about the Nix colour sensor.

(Refer Slide Time: 02:13)



And some of the keywords which we are going to discuss are Nix, organic carbon, colour space model, RGB and Munsell soil colour.

(Refer Slide Time: 02:24)



So, if you see the portable XRF and diffuse reflectance spectroscopy, we have already seen that these two methods are useful individually for predicting different soil properties. We have seen the application of DRS in our previous week. And in this week, we are talking about PXRF application for different soil properties. Although these two sensors are accurate, and they show huge potential for predicting multitude of soil properties.

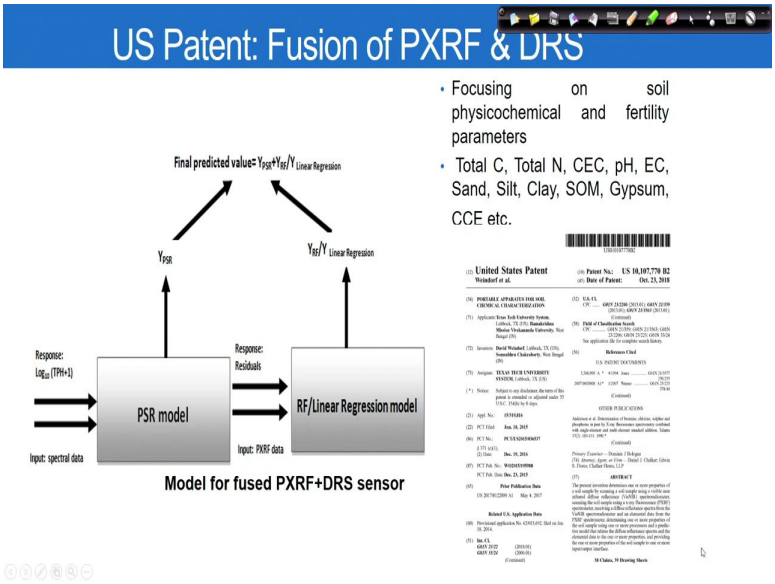
We have seen that when we combine the dataset from these two sources together, they gives some synergistic result. What is the synergistic result? As you can see in the topic of the slide that gives 2 plus 2 equal to 5 that means, when we combine the dataset from these two sensors together, we get better prediction accuracy than using the individual sensor in isolation. Because we have found there is some kind of complementarity between these two types of dataset.

Diffuse reflectance spectroscopy or in other words, the spectral method is sensitive to wearing colour properties or chromophores. For example, soil texture, then soil organic carbon, they impart differences in soil colour. So, this diffuse reflectance spectroscopy is sensitive to this colour change. However, the heavy metals do not have any spectral, direct spectral signature in the diffuse reflectance spectra, although they can be coexist, they can coexist with organic matter and this co-variation is helpful for their identification or prediction using diffuse reflectance spectroscopy.

On the other hand, portable XRF is not very suitable for predicting organic matter, but it can predict different metals and elements. So, when we combine the spectral data as well as PXRF data together, then we can see some kind of complementarity between these two types

of dataset. And as a result, they will give higher prediction accuracy then using individual sensors alone.

(Refer Slide Time: 05:03)



So, that notion of these synergistic effect by fusing the portable XRF as well as DRS together helps to generate the first patent in this aspect. Myself and Professor David Weindorf, we jointly applied for a US patent which shows the novelty of using PXRF and DRS data for better soil prediction. And so, we got this patent, then the name of this patent is portable apparatus for soil chemical characterization.

(Refer Slide Time: 05:49)

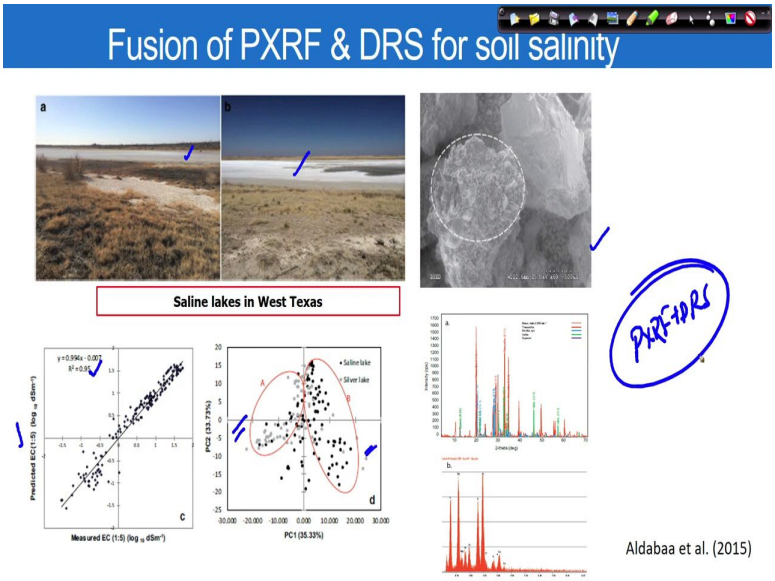
Now, remember, there are different ways through which we can fuse the dataset. One of the simple methods of using the dataset is simple concatenation. That means, we gave the

elemental data and we get the elemental data from PXRF, and we get the spectral data from DRS and we just simply merge them together in a single spreadsheet and then use them as predictors in the prediction model.

So, this is the simplest way of merging these two datasets. So, one of the first application we tried this sensor fusion approach is to predict the soil salinity. So, we collected several hundreds of soil samples from Western Texas of US and from some saline lake. And these saline samples as you can see, the salt crust are clearly visible on the soil, in the soil and these salt affected soils were first sampled and then scanned via PXRF as well as the diffuse reflectance spectroscopy.

Apart from combining these PXRF and DRS, we have also combined the Landsat bands, which is a satellite band, satellite sensor. So, Landsat satellite sensor bands are also used with portable experiment DRS data and all these three in combination produce higher prediction accuracy together for predicting the soil salinity.

(Refer Slide Time: 07:45)

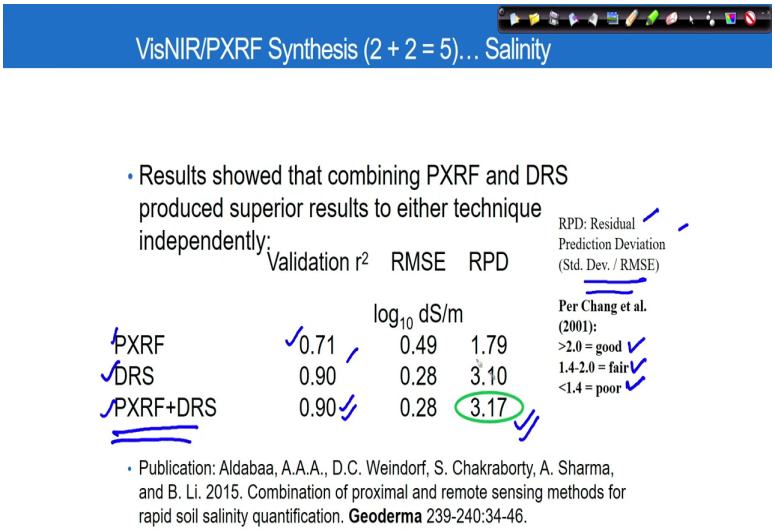


So, you can see these are the two location. We have collected the samples and using the support vector regression while combining all these data sets, we have got very good r square values with a value of 0.95. So, our model was highly accurate to predict the log transform electrical conductivity values as a measure of soil salinity.

And using the principal component analysis, we were able to segregate the samples coming from two different locations, because of their which may be due to the elemental variation between the two types of soil and we have validated the results using scanning electron

microscopy as well as using the same EDS technology. And also, so we have proved the usefulness of the PXRF and diffuse reflectance spectroscopy method for prediction of soil salinity.

(Refer Slide Time: 09:03)

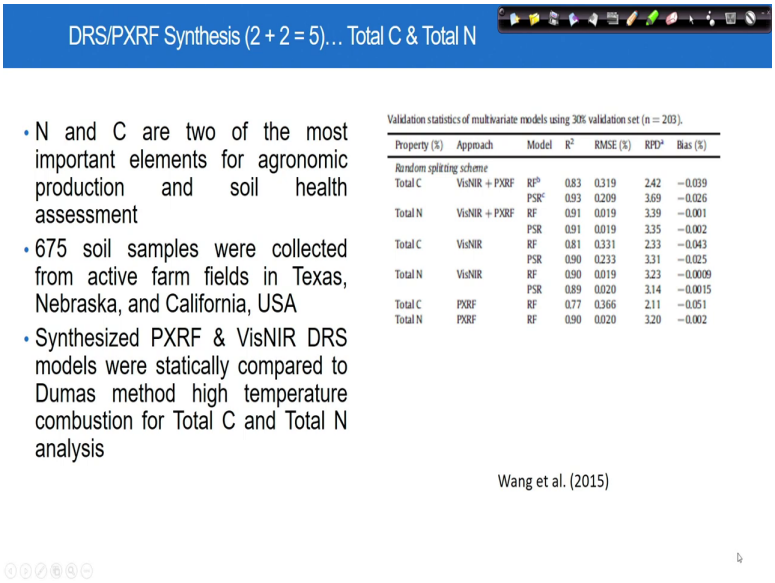


So, if you can see that results, where we are using the PXRF alone, we are getting an r square validation, r square value of point r square, validation r square values of 0.71. And when we are using the diffuse reflectance spectroscopy alone, we are getting 0.90 when also when we are combining the PXRF plus DRS although we are getting the similar 0.90 r square, but if you compare the RPD values, RPD stands for residual prediction deviation, which is basically the standard deviation minus RMSE of the model.

So, if you see specifically, we calculate this for validation samples. So, the validation standard deviation by the RMSE we get the RPD value. So, according to Chang et al if we get the RPD values of less than 1.4 that shows poor model performance, if the RPD values lies between 1.4 to 2 that shows fair model performance and if it is more than 2 that shows good and reliable model performance.

You can see that when we are combining the PXRF and DRS data, we are getting 3.17 which is highest among all these three models. The model which I showed you the SVR model, which is showing the r square values of 0.95 that is basically when we are combining PXRF DRS and also the remote sensing data and this is basically a combination of PXRF and DRS data. So, you can see that when we are combining the PXRF and DRS data, we are getting high RPD values, which shows the importance of the model fusion or sensor fusion.

(Refer Slide Time: 11:04)

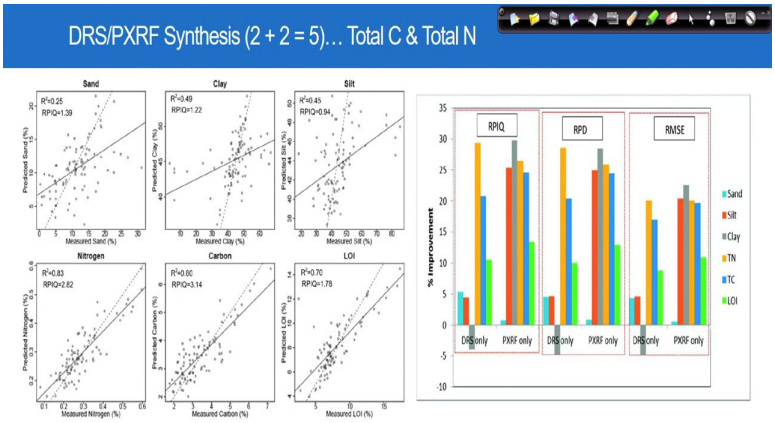


So, also in subsequent research, we try to model the total carbon and total nitrogen, these two are important soil parameters by combined diffuse reflectance spectroscopy and PXRF model. So, these two properties were modelled using 675 soil samples collected from Texas, Nebraska, California, of United States. And we have seen that the combined PXRF and DRS model where status were better, then using these PXRF and DRS individually or in isolation.

So, that shows the again the importance of PXRF as well as DRS combination through sensor fusion. And it was also published in a very good journal called Geoderma. You can see here the total carbon was, and both total carbon and total nitrogen were predicted using the combined DRS and PXRF using two models. One is random forest and other is penalized spline regression.

We will discuss this penalized spline regression in our upcoming lectures, which is a spline-based regression we will discuss this, but remember that in this research, we have found that when we combine the PXRF and DRS data together, we can get better results than using the individual sensor alone. So, that shows the importance of sensor fusion or multi sensor fusion for producing the better soil prediction.

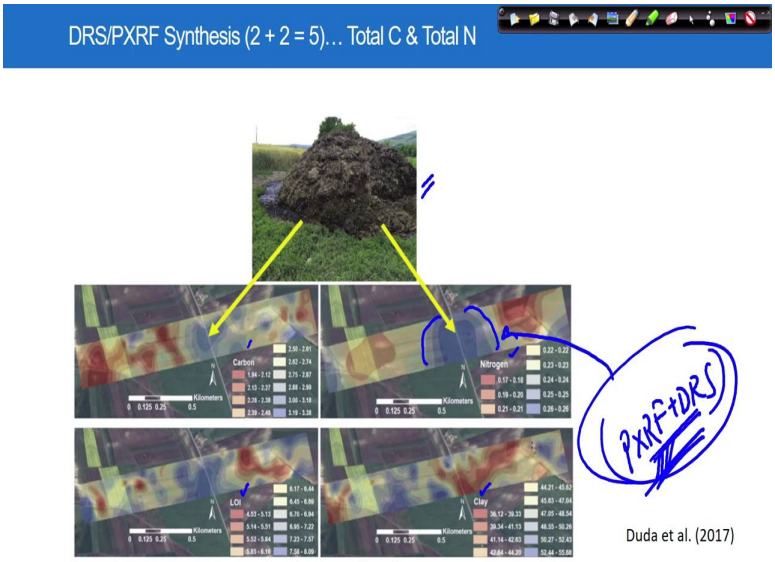
(Refer Slide Time: 12:56)



Duda et al. (2017)

Now, another research we executed in 2017 and it was also published in Geoderma. Here also we try to predict the total carbon and total nitrogen using combine diffuse reflectance spectroscopy and PXRF. And we have found that when we combine these diffuse reflectance spectroscopy and PXRF together, we got better model accuracy then predicting the total carbon and total nitrogen using individual sensors. We compared their RPIQ, RPD, RMSE, which are performance matrix and we have seen that combined model produce always better results than using the sensors in isolation.

(Refer Slide Time: 13:54)



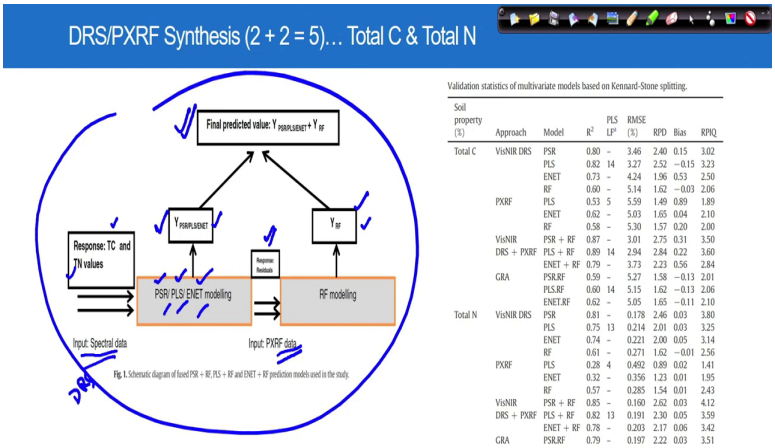
Duda et al. (2017)

So, you not only we have produced we have showed the importance of combined model but also at the same time, we produce the prediction map through kringing. So, these are the four

different maps you can see, total carbon, total nitrogen, then clay and loss on ignition, organic carbon. So, these maps were used, these maps were produced by combining these PXRF and plus DRS model, fused model and interpreted through Kriging interpolation.

So, if you can see that, the variation of nitrogen in this zone, so that shows that high content of nitrogen was found in this patch and we have seen that this is due to the presence of a pile of manure, so that shows that this combination of PXRF and DRS realistically identified these higher nutrient concentration in the proper space. So, that shows that not only this synthesized or fused model can produce better prediction accuracy, but at the same time it can produce better mapping of soil properties.

(Refer Slide Time: 15:31)



Validation statistics of multivariate models based on Kennard-Stone splitting.

Soil property (%)	Approach	Model	R^2	PLS LF[a]	RMSE (%)	RPD	Bias	RPIQ
Total C	VisNIR DRS	PSR	0.80	–	3.46	2.40	0.15	3.02
		PLS	0.82	14	3.27	2.52	−0.15	3.23
		ENET	0.73	–	4.24	1.96	0.53	2.50
		RF	0.60	–	5.14	1.62	−0.03	2.06
	PXRF	PLS	0.53	5	5.59	1.49	0.89	1.89
		ENET	0.62	–	5.03	1.65	0.04	2.10
		RF	0.58	–	5.30	1.57	0.20	2.00
	VisNIR	PSR + RF	0.87	–	3.01	2.75	0.31	3.50
	DRS + PXRF	PLS + RF	0.89	14	2.94	2.84	0.22	3.60
		ENET + RF	0.79	–	3.73	2.23	0.56	2.84
	GRA	PSR.RF	0.59	–	5.27	1.58	−0.13	2.01
		PLS.RF	0.60	14	5.15	1.62	−0.13	2.06
		ENET.RF	0.62	–	5.05	1.65	−0.11	2.10
Total N	VisNIR DRS	PSR	0.81	–	0.178	2.46	0.03	3.80
		PLS	0.75	13	0.214	2.01	0.03	3.25
		ENET	0.74	–	0.221	2.00	0.05	3.14
		RF	0.61	–	0.271	1.62	−0.01	2.56
	PXRF	PLS	0.28	4	0.492	0.89	0.02	1.41
		ENET	0.32	–	0.356	1.23	0.01	1.95
		RF	0.57	–	0.285	1.54	0.01	2.43
	VisNIR	PSR + RF	0.85	–	0.160	2.62	0.03	4.12
	DRS + PXRF	PLS + RF	0.82	13	0.191	2.30	0.05	3.59
		ENET + RF	0.78	–	0.203	2.17	0.06	3.42
	GRA	PSR.RF	0.79	–	0.197	2.22	0.03	3.51
		PLS.RF	0.64	13	0.262	1.68	0.02	2.65
		ENET.RF	0.65	–	0.256	1.71	0.03	2.70

[a] PLS LF, partial least squares latent factor.

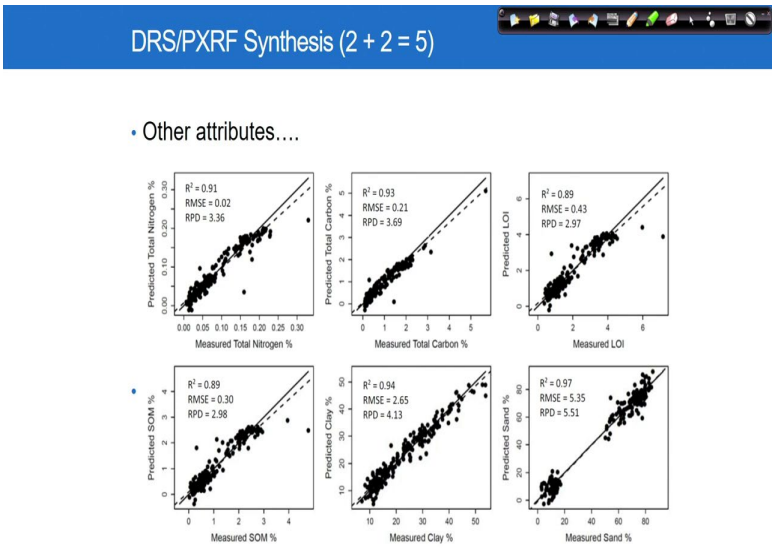
Cardelli et al. (2017)

Subsequently in another research, we tried to model the total carbon and total nitrogen. And here, instead of simple concatenating the data simple and merging the data, we have used another strategy for model fusion. As you can see here, here first we started with the total carbon and total nitrogen these were our response values or target values. So, in first instance our input was spectral data coming from DRS and using the spectral data we fit the model using penalized regression, partial least squares regression and elastic net regression.

So, we tried and compared three different models and then we predicted the total carbon and total nitrogen. Obviously, for any model there will be some residuals. So, the subsequently the residuals from all these three models were predicted or modelled using the PXRF data. So, simultaneously we are getting the predicted results from both these PSR or PLS or ENET model using the spectral data.

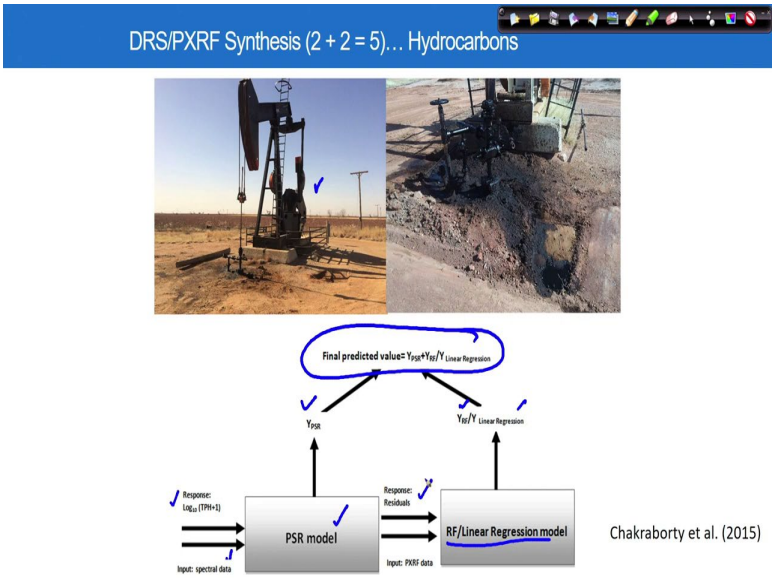
And also, simultaneously we are getting the prediction from this random forest model using the PXRF data. If we add then the final predicted values will be basically a combination of this component and this component. So, this is another way of fusing model and this new novel strategy was also used for predicting the total carbon and total nitrogen by Cardelli et al in 2017. And we have proved that using this strategy also, we can see that we can produce better model accuracy that using these individual sensors alone.

(Refer Slide Time: 17:40)



So, there are several other attributes also, which can be predicted using this sensor fusion approach. These are total nitrous and total carbon, I have already showed you then loss on ignition, soil organic matter, clay contents, sand content all of these attributes have showed good results and supported the use of sensor fusion for their better prediction accuracy.

(Refer Slide Time: 18:10)

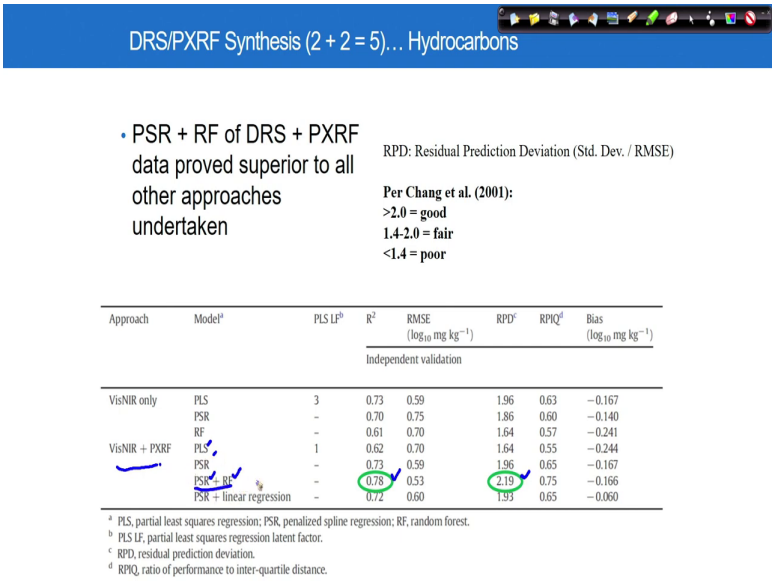


Another application we have tried is to predict the petroleum hydrocarbon. So, you can see that these petroleum pump jacks or the petroleum extraction process sometime contaminate the surrounding fields. So, this is a picture taken in a cotton field of Texas and we can see that this crude oil which are being extracted from the soil sometimes get spilled away and contaminate the surrounding land.

So, this petroleum contaminated soils should be predicted using some sense, using some advanced sensing methods. So, we tried to use this combined sensor method also to predict these petroleum hydrocarbon contamination. So, in this case, our response was longer than we converted this total petroleum hydrocarbon and we model it using the spectral data. So, you can see here we tried to model this logarithmic log converted the total petroleum hydrocarbon using the spectral data using penalized spline regression model.

And then, the residuals were further predicted using either random forest or linear regression model using PXRF data. So, we got Y PSR or we got Y RF or Y linear model regression, which had the predicted values and then the final predicted values is basically the combination of these two terms. We have seen that when we combine these two dataset, we combine these two dataset by using these model fusion technology, then we can get higher prediction accuracy.

(Refer Slide Time: 20:15)

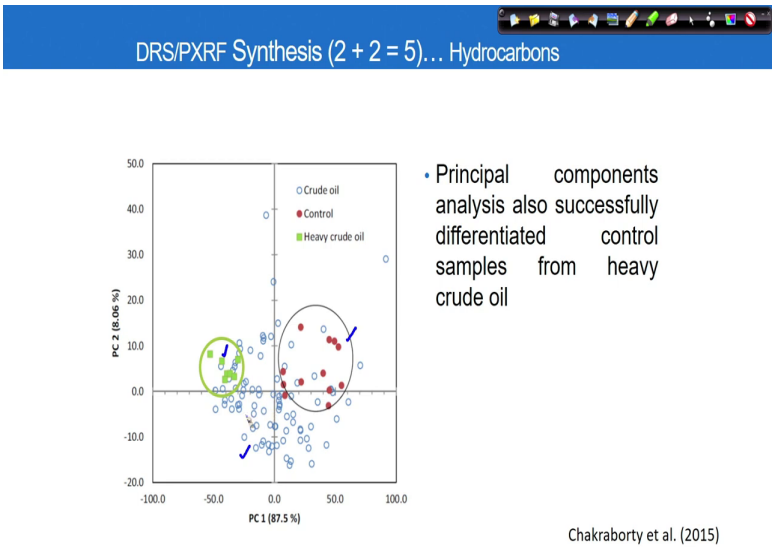


Approach	Model[a]	PLS LF[b]	R^2	RMSE (log_{10} mg kg^{-1})	RPD[c]	RPIQ[d]	Bias (log_{10} mg kg^{-1})
			Independent validation				
VisNIR only	PLS	3	0.73	0.59	1.96	0.63	−0.167
	PSR	–	0.70	0.75	1.86	0.60	−0.140
	RF	–	0.61	0.70	1.64	0.57	−0.241
VisNIR + PXRF	PLS	1	0.62	0.70	1.64	0.55	−0.244
	PSR	–	0.73	0.59	1.96	0.65	−0.167
	PSR + RF	–	0.78	0.53	2.19	0.75	−0.166
	PSR + linear regression	–	0.72	0.60	1.93	0.65	−0.060

[a] PLS, partial least squares regression; PSR, penalized spline regression; RF, random forest.
[b] PLS LF, partial least squares regression latent factor.
[c] RPD, residual prediction deviation.
[d] RPIQ, ratio of performance to inter-quartile distance.

As you can see from these results, when we are combining these penalized spline regression and random forest not only, we are getting the highest r square, but also, we are getting the highest RPD values among all the models, where we have tried these, either we have tried these individual DRS or other models, like PLS, PSR or linear regression combination. So, we have found that not only the sensor fusion is important, but also combination of model is also very important for producing the better results.

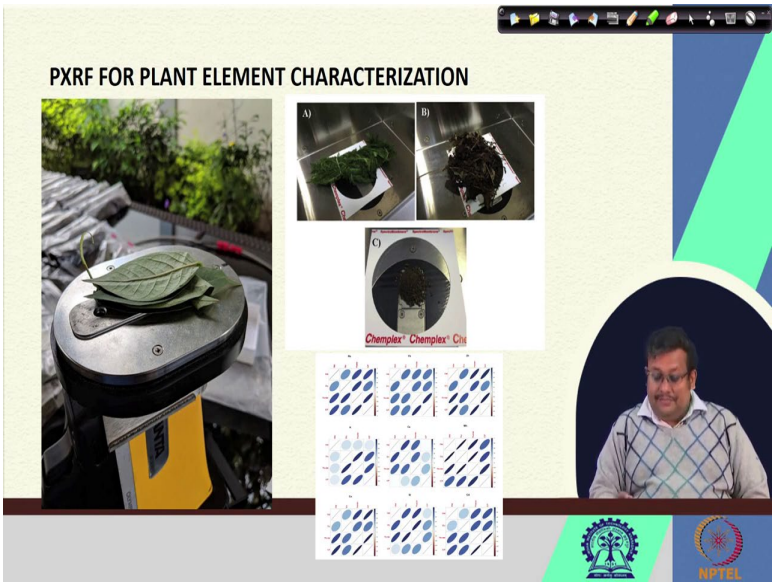
(Refer Slide Time: 21:02)



We have also seen that it is possible to segregate the soil samples or crude oil contaminated samples. So, here you can see that these green points are showing heavy crude oil contaminated sample, these red points are showing these control samples which are non-

contaminated and these blue circles are showing the crude oils contaminated samples, it is possible to segregate the samples based on their contamination level.

(Refer Slide Time: 21:40)



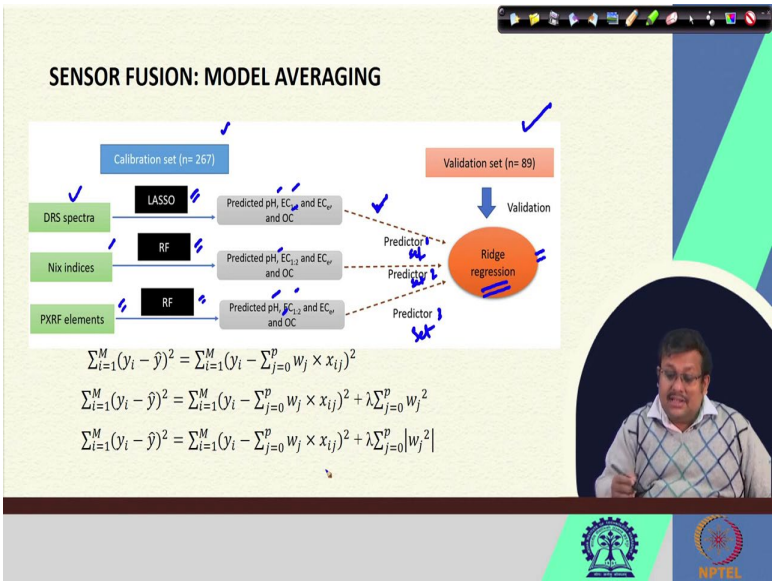
So, so far guys we have talked about the combined sensor platform. And also, we have talked about the model fusion. So, we move from there and also try to use the portable XRF for plant elemental characterization, you know that plant has different parts and different parts can accumulate different elements. So, the elemental content in the plant is also very important and generally we measured them in the lab after total digestion process which involves caustic chemicals like acids and also digestion process which is very hazardous which is very much exothermic.

So, to replace or to supplement those hazardous time consuming and corrosive chemical intensive elemental concentration measurement in plant bodies, we also use the portable XRF for predicting the elemental content, we have used different plant parts. And we have seen that PXRF is very much helpful for predicting accurate elemental content. So, we got very high correlation between the PXRF elemental content as well as the total elemental content which we get from the traditional methods.

So, that shows that this PXRF cannot be only used for soil, but also it can be used for other matrices like plant matrices. So, but at the same time, it is also advisable that since the commercially available PXRF are only having a limited number of internal calibrations focusing on geochemistry soil, it is important to expand this library for other matrices also like plant because it depends on metrics density.

So, since the plant metrics density is different than soil metric density, we need to develop individual and separate prediction model for plant materials and plant elements also and we should incorporate that calibration within our next generation of PXRF.

(Refer Slide Time: 24:09)



And other sensor fusion technology we have tried that is called model averaging. And in this model averaging you can see we try to use different models. For example, here you can see that a whole dataset is segregated into calibration set and validation set and using the calibration set. Suppose, we fitted this lasso regression, we will be discussing this lasso regression in our upcoming lectures.

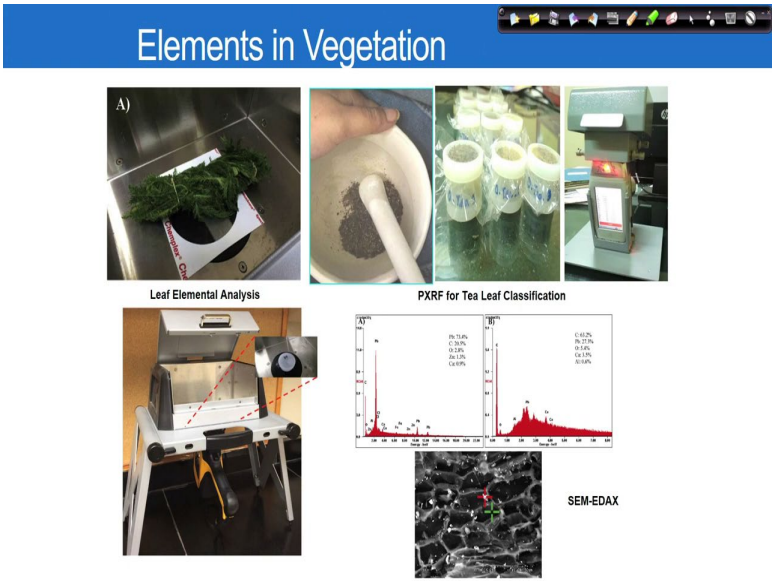
Remember that this DRS spectra was used to predict the pH, EC and organic carbon of the soil using the lasso regression. The Nix which we are going to discuss, Nix is the colour sensor. So, Nix indices colour indices are also being used for predicting these same parameters using random forest model and PXRF elements were also used for predicting the pH, EC and mostly using random forest model. So, for all these three models, we can get the predicted values for this calibration set.

So, predictor 1, predictor 2 and predictor 3 or predictor set 1, predictor set 2 and predictor set 3. So, all these were combined to use the final model through ridge regression, we are also going to discuss this ridge regression in our upcoming lectures. So, these predicted calibration samples are our predicted values of the calibration samples were used as predictor sets for and then they were subsequently combined to produce the final prediction model through ridge regression. And that model was validated using the validation set. So, this is called model averaging.

This model averaging has been used in other literature in other applications also, some soil scientists have used different versions of model averaging like Granger and Ramanathan model averaging method they have tried for combination of PXRF as well as DRS data. So, there are different approaches for combination of dataset, remember, it depends upon your particular application, sometimes you will see that simple concatenating or simple merging of these two dataset is better and sometimes you have to use these either model averaging or these model fusion strategies for getting the better results.

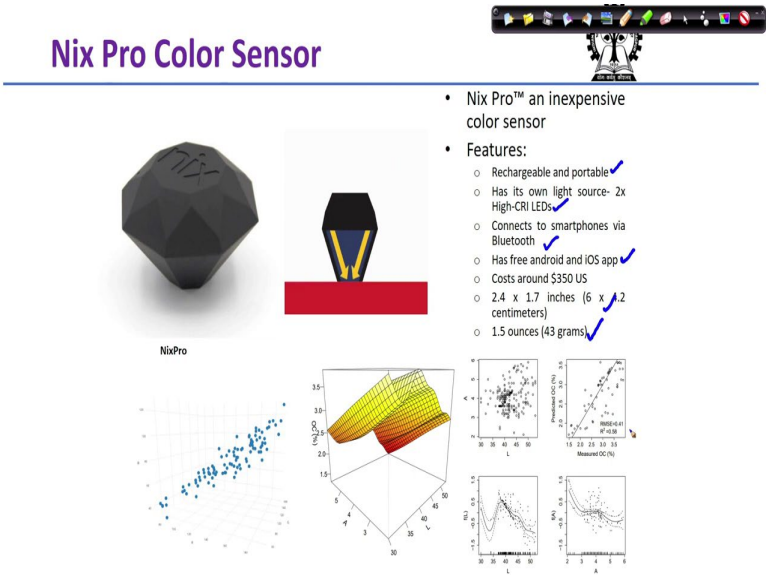
And there is no universal best model, you have to develop this model and you have to try different combinations that you have to you have to try different types of model to see which one works better with your dataset. And then you have to select that for subsequent analysis.

(Refer Slide Time: 27:28)



We have also seen that PXRF can be also, as I have told you that PXRF can be used for elemental content analysis in the vegetation, we have extensively used these for heavy metal content identification, heavy metal identification in tea leaves, and based on this elemental content of these tea leaves, we try to classify these tea leaves using different types of classification schemes.

(Refer Slide Time: 27:55)

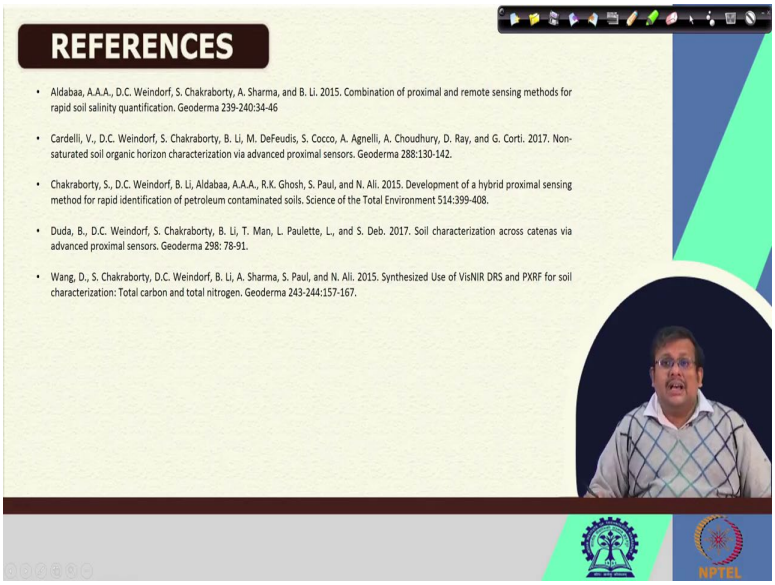


So, another sensor we have used is Nix Pro Colour Sensor or Nix Colour Sensor, which is an inexpensive colour sensor. Remember this is a very recently used colour sensor for soil and soil colour is defined by using the Munsell soil colour chart, which is qualitative definition. So, these Nix probe gives you the quantitative results in terms of different types of colour models like RGB, CMYK, CLAB, LCH and so on.

So, it is a, Nix Pro is an inexpensive colour sensor, it is relatively cheap, add it is rechargeable, and portable. It has its own LED light source. It can connect to the smartphone and operated through smartphone via Bluetooth and smartphone Android app and iOS app and then it is very small and also very lightweight also, only 43 grams.

So, this sensor has been used for last 4 to 5 years in different domains of soil and we are going to discuss that, but remember that this is now, this sensor has become an important topic of discussion among the soil scientist. Nowadays, for better prediction of several soil properties. We are going to discuss the application of Nix Pro Colour Sensor or Nix Colour Sensor in our upcoming slides.

(Refer Slide Time: 29:42)



So, guys, let us finish this wrap up this lecture here. These are the references for this lecture. And in our next lecture we are going to talk more about the Nix Colour Sensor and their application. And we will be seeing how Nix can be also combined with other sensor to produce better results and how the other sensors, crop sensors can be used along with machine learning tools for predicting different types of crop properties. So, let us wrap up this lecture here. Thank you and let us meet in our last lecture of week 6.