Machine Learning for Soil and Crop Management
Professor Somsubhra Chakraborty
Agriculture and Food Engineering Department
Indian Institute of Technology Kharagpur
Lecture 23
Diffuse Reflectance Spectroscopy: Basics and Applications for Crop and Soil (Contd.)

(Refer Slide Time: 00:25)

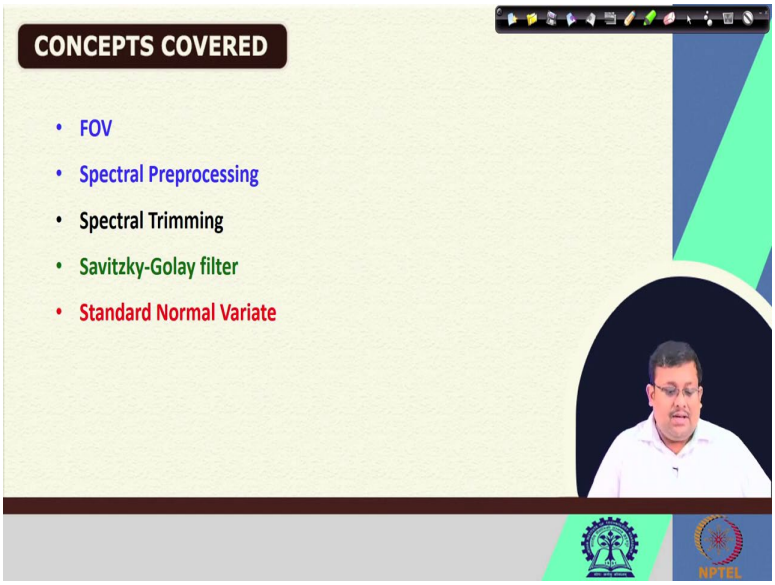
Welcome friends to this 3rd lecture of week 5 of NPTEL online certification course on Machine Learning for Soil and Crop Management. And in this week, we are discussing about Diffuse Reflectance Spectroscopy and Basics and Application for Crop and Soil. This is lecture number 23, our 3rd lecture of this week 5.

And in our previous two lectures, we have discussed some of the important concepts, we have discussed the multispectral remote sensing, hyperspectral remote sensing, what are the differences, what is soil spectroscopy and what is the difference between mid-infrared spectroscopy as well as near visible to near infrared spectroscopy, why mid infrared spectroscopy contains more information than visible near infrared spectroscopy we have discussed.

So, and also we have discussed different types of specifications of the visible near infrared spectroscopy. We have seen what is sampling interval, what is spectral resolution we have discussed. And also, we have seen the working principle of different types of detectors like VNIR detector, which is 512 silicon photodiode array, and indium, gallium arsenide detector, which is used for is SWIR1 and SWIR2 region we have seen.

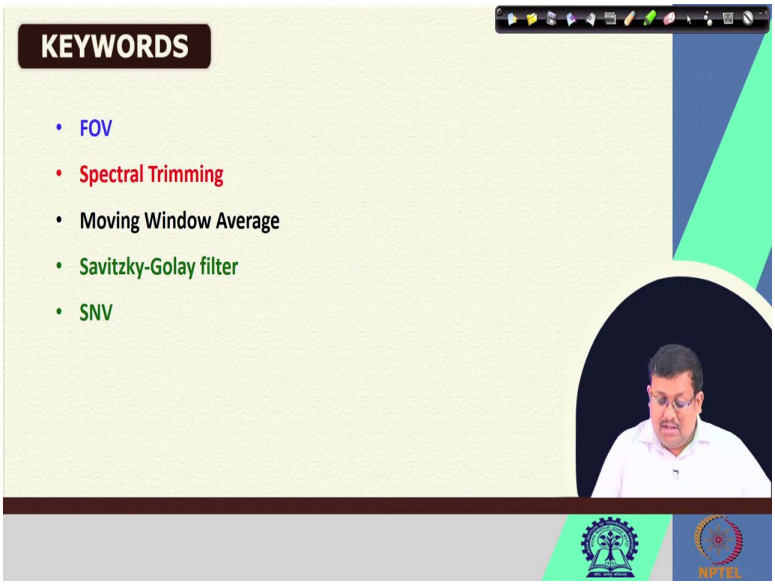
So, today we are going to start from there and we are going to discuss some of the other important specs like FOV, what is field of view and then we are also going to start a very important aspect that is spectral prepossessing.

(Refer Slide Time: 02:19)



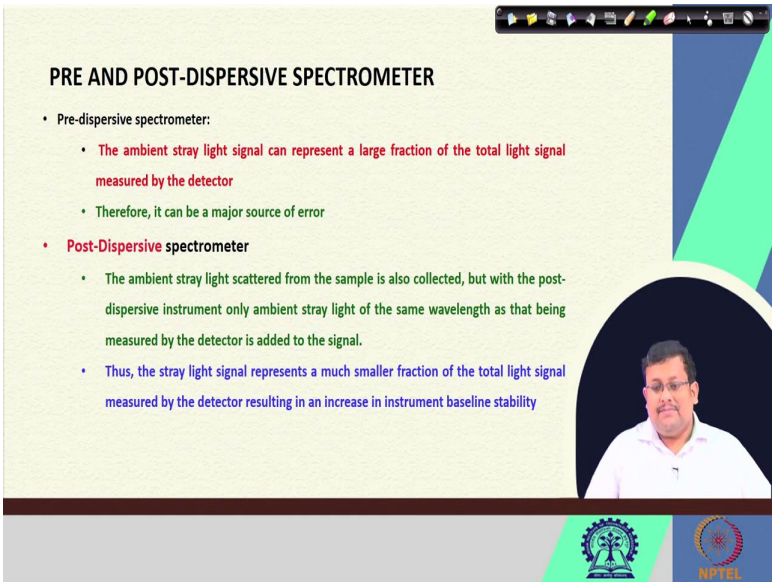
So, these are the topics which we are going to cover in today's lecture one is the FOV or field of view and then we are going to discuss the spectral prepossessing. Some of the important spectral prepossessing like spectral trimming, Savitzky-Golay filter, standard normal variate, detrending we are going to discuss. So, most of the discussion in today's lecture in this lecture will be focused on the spectral prepossessing.

(Refer Slide Time: 02:49)



So, these are the keywords for today's lecture for this lecture, FOV, spectral trimming, moving window average and then Savitzky-Golay filter, SNV which we are going to discuss.

(Refer Slide Time: 03:03)

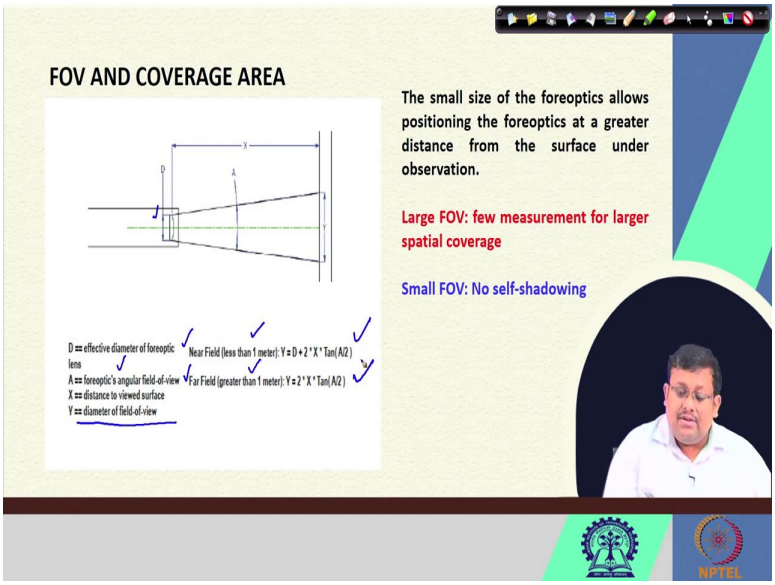


So, in the previous lecture, we have stopped here that is discussing about the pre and post dispersive spectrometer. So, what is the difference between this pre and post dispersive spectrometer? There are two types of spectrometers one is pre-dispersive another is post-dispersive. Now, in case a pre-dispersive spectrometer, the ambient light that means stray light signal can represent a large fraction of the total light signal measured by the detector.

So, as a result of that, it can be a major source of error. So, that creates a problem in case of user pre-dispersive spectrometer. And other one spectrometer is there that is called post-dispersive spectrometer. So, here the ambient straylight scattered from the sample is also collected, but with the post-dispersive instrument only ambient stray light of the same wavelength as then being measured by the detector is added to the signal.

So, this stray light signal represents a much smaller fraction of the total light signal measured by the detector resulting in instrument baseline stability. So, that is why the pre-dispersive spectrometer is not preferred over the post-dispersive spectrometer. Post-dispersive spectrometer is rather preferable than the pre-dispersive spectrometer.

(Refer Slide Time: 04:28)



Another important concept is a field of view that means using different types of foreoptics like we have different types of lens, we have different types of probe. How much area we can cover? What is the field of view? So, this selection of this field of view generally varies based on our objective. So, and the selection of the field of view should be done very carefully because they are some pros and cons of both small and large field of view.

So, generally in case of the spectroradiometers which are commercially available, they are having very small size foreoptics, we are going to discuss like contact probe, this contact probes are very small size. So, you can position this contact probe of this foreoptics at a greater distance from the surface under observation, we will see some photographs.

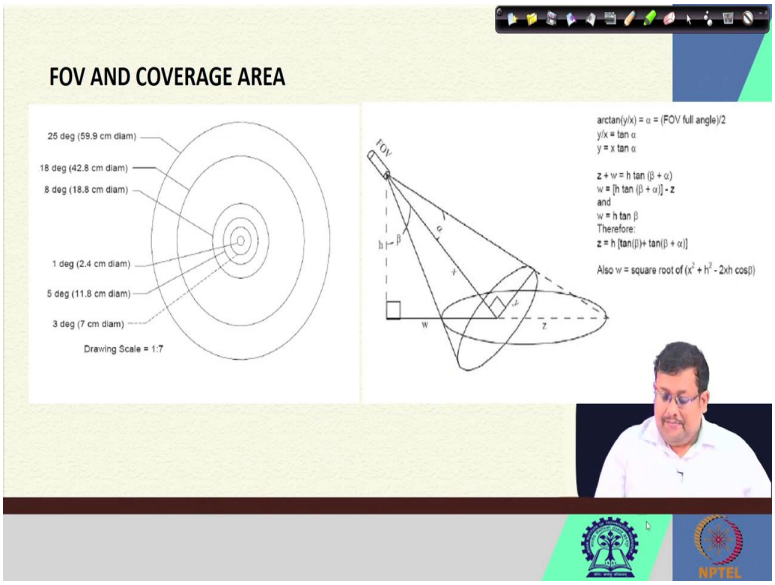
So, but at the same time, you should remember that, in case of large FOV when we are talking about large FOV, that larger FOV generally show few measurement for larger special coverage. So, when you have to have a large special coverage, we generally go with the large field of view, but when there is a small FOV that helps in removing the self-shadowing effect.

So, this is the difference between small FOV and larger FOV, but in case of large FOV you can have a special coverage. So, you can have, you can do a very few number of measurements to cover a large area. So, this is how you can calculate the effective diameter of, the diameter of the field of view or Y, you can see here the formula is given. So, a diameter field of view which is denoted by Y is given.

So, in case of near field which is less than 1 nanometer, 1 meter and there is a far field which is greater than 1 meter distance. So, in case of nearfield the formula is D plus 2 X Tan A by 2 where D is the effective diameter of the foreoptics lens. So, this is the effective diameter of the foreoptics lens, this is the foreoptics and this is the effective diameter of the foreoptics lens. X is the distance of the viewed surface, so that is called as X.

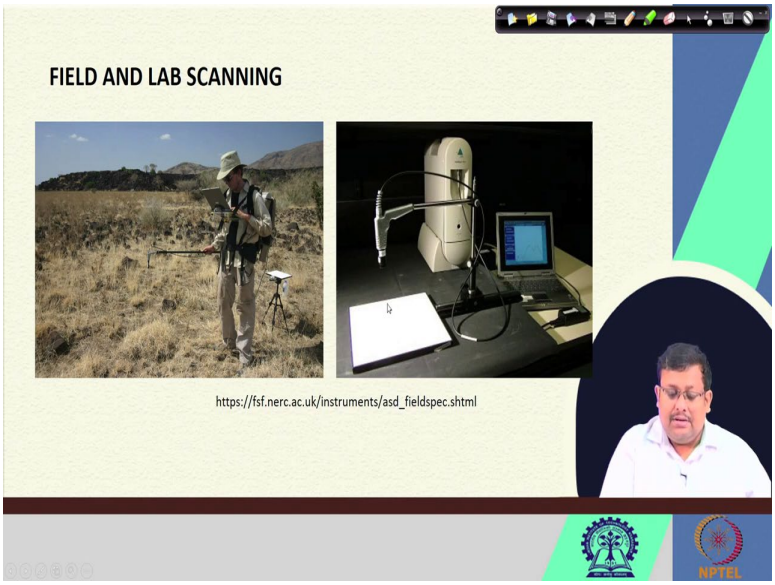
And A is basically the foreoptics angular field of view, so it is the angular field of view. So, this is how you calculate the diameter of field of view, use for the near field condition. But in case of far field condition when it is greater than 1 meter, then you utilize this formula that is 2 X Tan A by 2. So, 2 X, so this is the formula for both the near field FOV and the far field FOV calculations. So, this is how you calculate the field of view. So, the large FOV is useful for special information, small FOV helps in no self-shadowing.

(Refer Slide Time: 07:53)



So, you can see here that as we move from different types of FOV, this is the 1 degree, then 5 degree, then 3 degree, 5 degree and as we go to the 25 degree of course, the diameter is increasing. So, higher field of view is showing the, higher field of view is covering the larger area than that of the smaller field of view. And also, for this tilted orientation, how to calculate the FOV is also the formula is given here. So, you can utilize this formula to calculate the diameter of field of view or the area of the field of view you can calculate.

(Refer Slide Time: 08:52)



Now, of course, this as I have told you, that this spectroradiometer can be utilized both in field and lab and as you can see, you can take a field measurement by putting these foreoptics at a certain distance from the surface, you can see these are very small. So, it is having a small FOV and the larger FOVs are also there, but at the same time you can see whenever we are doing this optical measurement, there are a wide spectrum and panels are always there for referencing your spectroradiometer.

And also, this is the fields this is a laboratory setup, which you can use for making the laboratory measurement. This is the fore-optic, it is connected to the spectroradiometer and this is the wide spec (())(09:49) panel. You can take the scan of the any surface just keeping the sample below in any powder form.

(Refer Slide Time: 10:00)

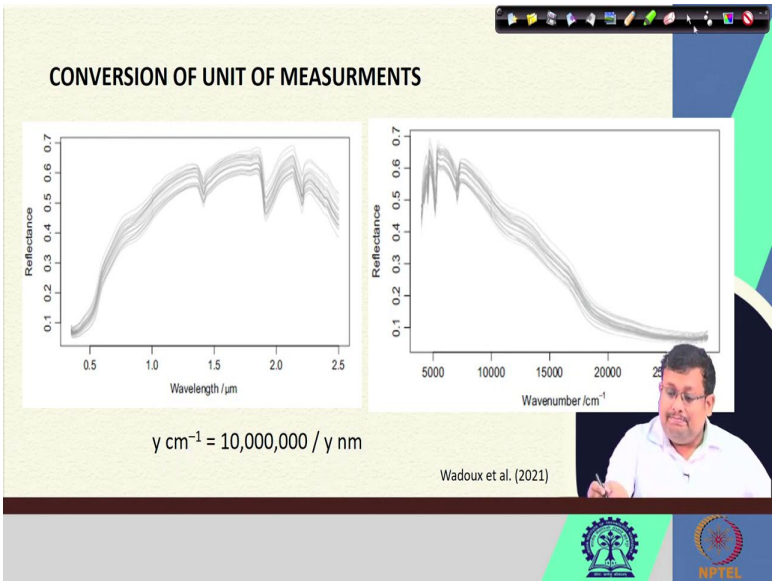
Now, there are different types of foreoptics, these are generally I have given here some example of the foreoptics, which are there with the field spectrometers, you can see here the contact probe, which is used for directly touching the powder surface and then collecting the spectral response. And this is a high bright contact probe, which is useful, which is also similar to this contact probe, but it has a high intensity light source.

And these are called the Hi-Brite Muglight. So, whenever we go for the field scanning, lab scanning, we can keep the sample just over these Muglight and we can scan it from below. So, generally the idea is we take it especially optically transparent petri dish. And this optically transparent petri dish we keep this sample in it and we place over this Muglight and it scan from below.

So, this is another arrangement for doing the laboratory measurement. This is for plant probe and also heat the plants leaves are heat sensitive. So, special care has been taken by producing these plant probe which are for the heat sensitive surface and this is useful for scanning the plant bodies or plant leaves.

Now, also this called a leaf clip which is useful for scanning the leaf and also this called a pistol grips. This pistol grips are useful for using the foreoptics directly in the field. These are very handy and you can use very easily to use the foreoptics directly in the field specifically in the in case of field measurement.

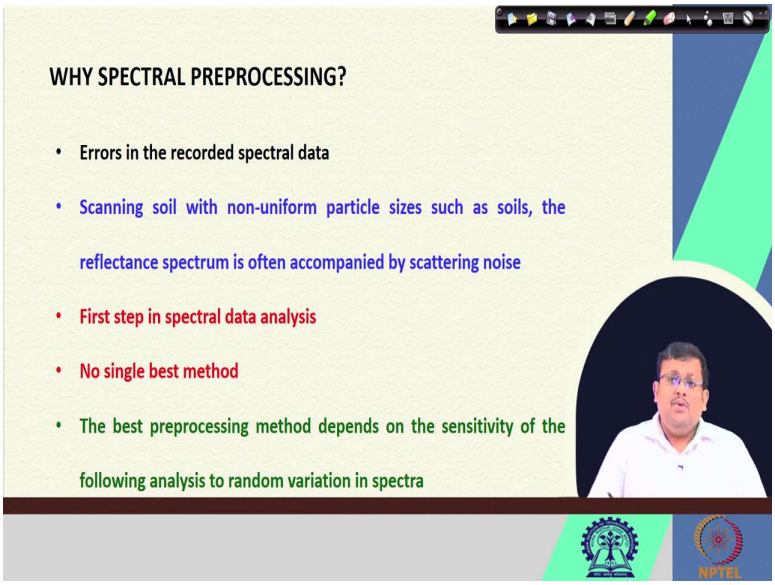
(Refer Slide Time: 12:00)



So, we have discussed the some of the important specifications of the diffuse reflectance spectroscopy. Now, we are going to start the discussion about the spectral prepossessing. Now, generally you can see here that we can see, we can plot the spectra using different types of unit of measurement. Most generally, we use the it this wavelength in terms of micrometer and nanometer. The reflectance values, the reflectance values are plotted here in terms of micro, in terms of wavelength in micrometer.

And also, we can represent or we can plot the same reflectance values in terms of wave number which is centimeter inverse. So, this is the formula of calculating the wave number from any given wavelength in nanometer. So, once we have a nanometer or micrometer, we can convert it to the wave number and we can plot. So, it is basically simple conversion the unit of measurement for this reflectance spectra.

(Refer Slide Time: 13:19)



Now, the question comes why we go for the spectral prepossessing, because as you know that while we capture the spectra specifically in the field or sometime in the lab, there are some errors in the recorded spectral data because it is spectral instruments, optical instruments are highly sensitive to the surrounding variation. So, when there is a variation in the surrounding specifically light condition, temperature condition that creates some variation in the spectrum.

So, scanning soil with non-uniform also, scanning the soil which is highly heterogeneous with non-uniform particle size. So, the reflectance spectrum is often accompanied by scattering noise. So, because as you know that soil contains sand, silt and clay which are having different size. So, when we take the scan will see the scattering effect. The noise scattering noise in the spectrum because of variation of the size of the particles.

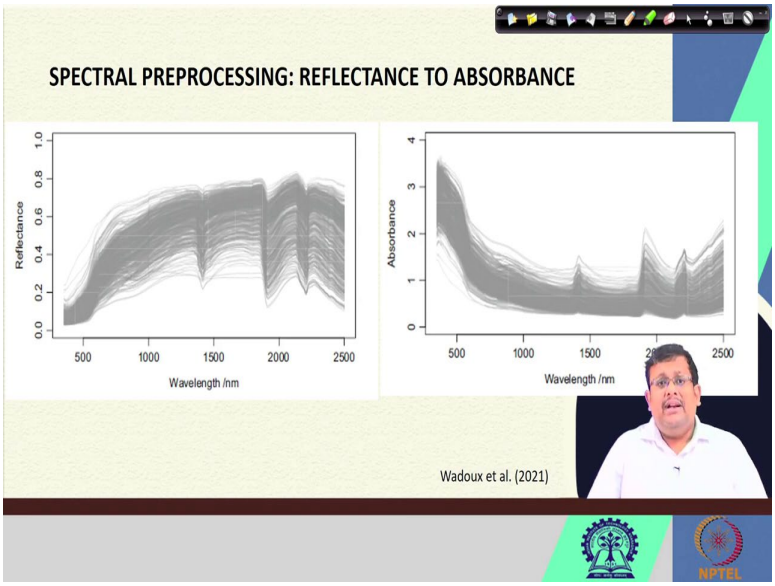
So, generally you remember that in case of spectro transfer modelling or spectral transfer modeling, the first step is the spectral prepossessing. If you want to look qualitatively discriminate the spectral also, you need to go for some kind of spectral prepossessing. So, spectral prepossessing is generally considered as the first step in spectral data analysis.

At the same time, it should be mentioned that there is no single best method and the best optimum method is to be selected by trial-and-error method, so the best spectral prepossessing method depends on the sensitivity of the following analysis to random variation in soil spectra. So, if the following method generally the quantitative method or the generally we call it a chemometric algorithm, they can identify the random variation in the spectra, then we can identify this is the best prepossessing method.

So, the best prepossessing method should be selected by in a trial-and-error basis and there is no uniform or optimal or globally optimal I would say spectral prepossessing method. There are various types of spectral prepossessing methods and depending on the data set, we have seen that for individual data set, specific a type of spectral prepossessing works better.

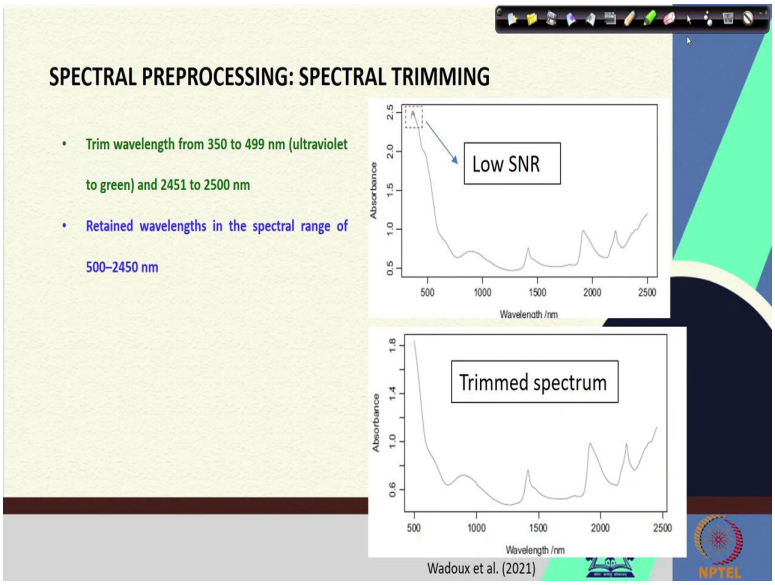
In some database, we have seen that first derivative of reflectance spectra works better, sometime we have seen only the smoothing of the spectra will work better, sometime we have seen standard normal variate works better, sometime we have seen multiplicative scatter correction works better. So, the suitability of a specific spectral prepossessing varies from one data set to another data set.

(Refer Slide Time: 16:27)



So, you can see here one example of spectral prepossessing is given here, this is reflectance to absorbance. So, we have the reflectance data we can convert it to the absorbance data also. So, this is one type of spectral processing and we have seen that instead of reflecting sometime the use of absorbance data also can produce higher, comparatively higher prediction accuracy in the subsequent calibration model. So, this is one of the important spectral conversion I would say.

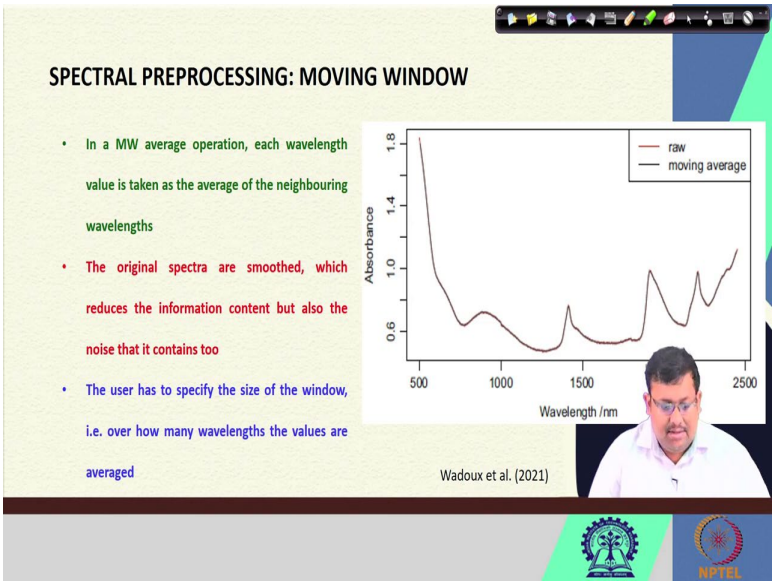
(Refer Slide Time: 17:07)



Now, another important spectral processing is called the spectral trimming. So, here you can see this is a raw-spectra and you can see here there are certain regions in generally, generally the certain regions in case of VisNIR spectrum appears at around 350 to 499 nanometer and from 2541 to 2500 nanometer. So, here you have very low SNR, signal to noise ratio. So, this signal to noise ratio does not, this low SNR region does not contain any useful information for the soil because they are present in the border lines.

So, we remove these low SNR regions to capture or to detain only this 500 to 2450 nanometer range which is having relatively high SNR then the stream region. So, this is one of the recent, one of the popular methods of spectral prepossessing. You trim the spectra and you keep the spectra from 500 to 2450 nanometer which and then you remove the two sides of the spectra which is having. And basically, when there is a low SNR that means, it is corrupted by noise. So, we remove these two regions from these two extreme sites to keep only the 500 to 2450 nanometer range.

(Refer Slide Time: 18:52)

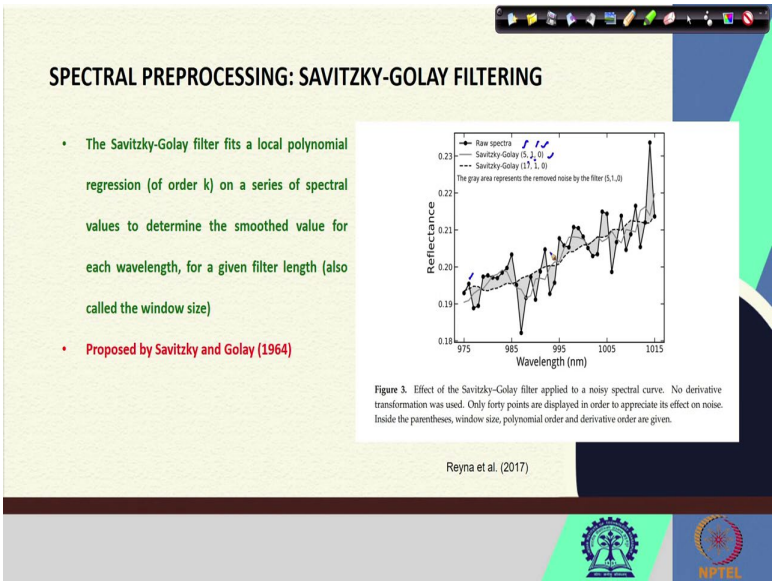


So, another spectral prepossessing method is known as the moving window average method. So, in moving into average operation each wavelength value is taken as the average of the neighbouring wavelength. So, the original spectra are smooth, which reduces the information content but also the noise that it contains too.

So, basically the idea is when you do the moving average window, sorry moving window averaging, then you reduce the noise but at the same time, you are losing some information also. So, the user has to specify the size of the window and over how many wavelengths values are average. So, basically it is an averaging and in moving basically each wavelength value is taken as the average of the neighbouring wavelength.

So, as it moves it takes the average and type then in kind of smooth the spectrum. Smoothing the spectrum is it is reducing the noise. And although it is reducing the noise, but at the same time it is using some of the important information also. So, we should be carefully or objectively select the applicability of this moving window averaging but this is a useful method.

(Refer Slide Time: 20:16)



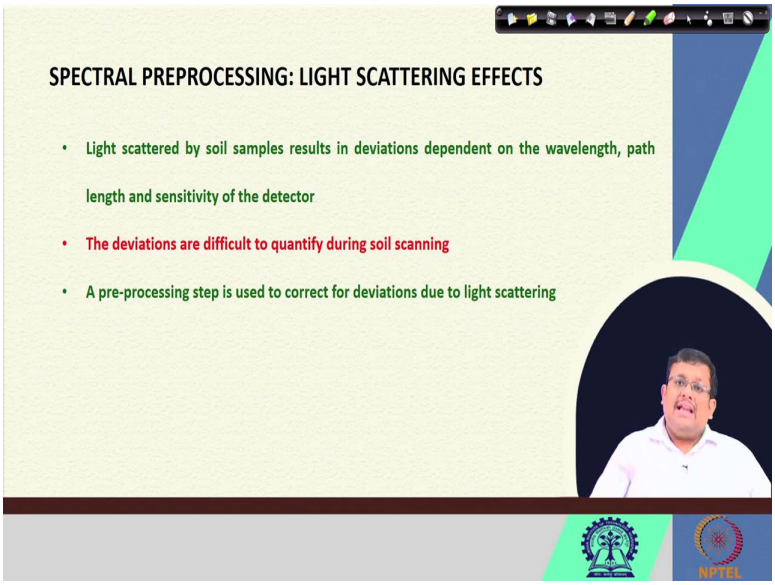
Now, let us see one of the most important spectral pre-processing method, which is called Savitzky-Golay filtering method. So, you can see this Savitzky-Golay Filter fits a local polynomial regression of order k on a series of spectral values to determine the smooth value for each wavelength for a given filter length. And this filter length is also known as the window size and this method of filtering, spectral filtering is given by proposed by Savitzky and Golay in the year 1964.

So, let us see one graphic or one example here from this plot. So, here you can see this is the raw spectra and then this is the Savitzky-Golay, Savitzky-Golay filtered spectra (())(21:11) smooth spectra, and this dashed line represents the another Savitzky-Golay spectra, but the difference between these two Savitzky-Golay filters spectra is given in the parenthesis which basically shows the window size, so here the window size is 5 here, window size is 17.

And also, the polynomial order is both the cases is 1. So, and also here the derivative order at 0, so there is no derivative, it is only the smoothing. So, you can see here based on the window size also how the smoothing will change the shape of the spectral curve. So, in case of when we are taking the higher window size, window size it is making the, it is losing huge amount of information and making almost these variations spectral, variations are almost not there.

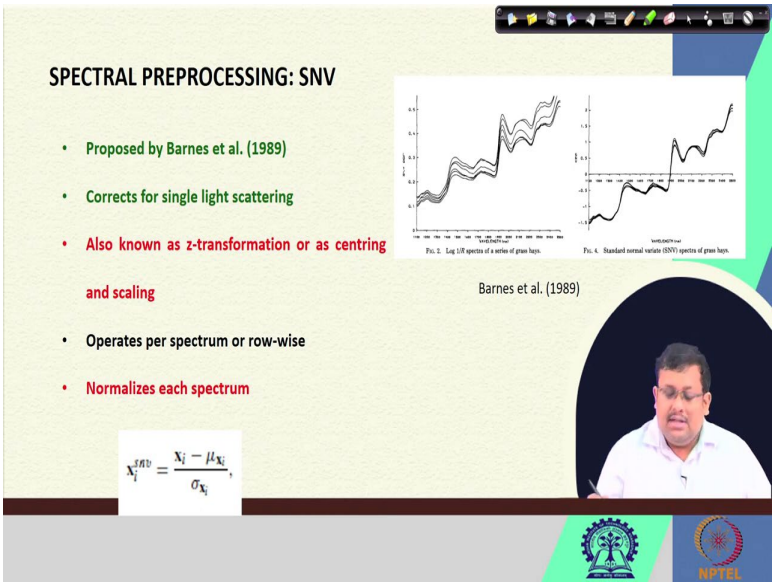
But when we are taking the relatively less filter length with which is only 5, then the window size 5 then we can see it is retaining the spectral variations. So, depending on your application you have to select which one you will select and you have to objectively select the best one out of the several options of window size.

(Refer Slide Time: 22:42)



Now, as I have told you that variation of the soil particle size can help, it can produce light scattering and it depends on the wavelength and also path length and also sensitivity of the detector. So, these deviations are difficult to quantify using during the soil scanning. So, a pre-processing step is used to correct for deviation, but due to the light scattering. So, this light scattering effect needs to be corrected. So, there are several methods like standard normal variate, multiplicative scatter correction, these are important spectral pre-processing generally scientists are following for removing these light scattering effects.

(Refer Slide Time: 23:34)



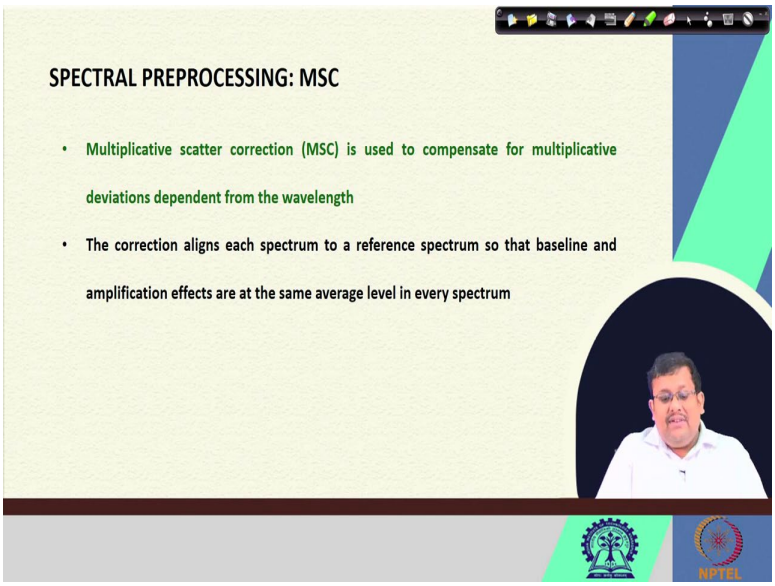
So, let us start with the standard normal variate. This standard normal variate was first proposed by Barnes et al. in the year 1989. And it basically helps in the correcting for single

light scattering. So, also this is known as the z-transformation or z-transformation or a centring or scaling. So, here you can see that these are the raw spectra and these are the SNV corrected spectra.

So, these are grass hays and original spectrum the grass hays and this is an absorbance spectra actually log 1 by R is an absorbance spectra and this is the standard normal variate spectra of these grass hays. So, generally this SNV works for individual spectrum or row size and it normalizes the spectrum by using this formula.

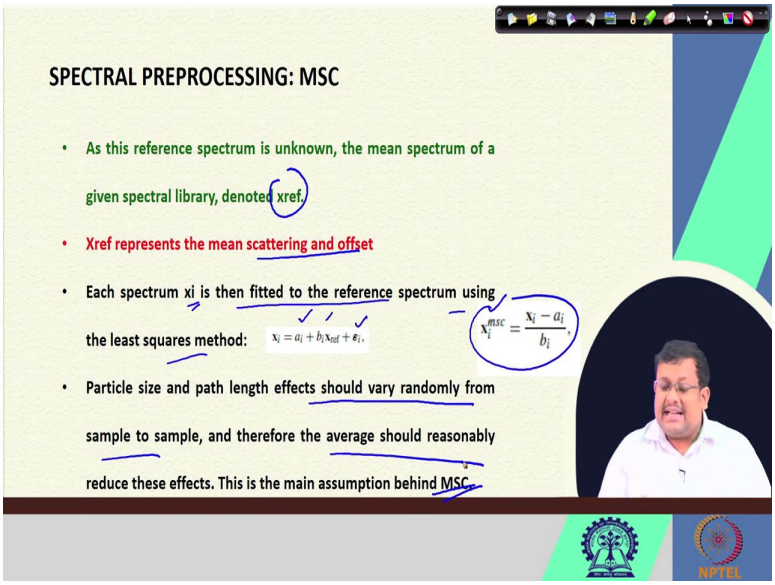
This Xi is the individual spectrum, the spectral values and then you subtract the mean and divide it with the unit variance. So, mean is 0, variance unit variants. So, you get the normalize each spectrum and then you get this type of shape. So, when you normalize the spectrum you can remove the light scattering effect.

(Refer Slide Time: 25:06)



Another one spectral processing pre-processing is multiplicative scatter correction. So, this multiplicative scatter correction is used to compensate for multiplicative deviation dependent from the wavelength or deviations which are dependent on the wavelength. So, the correction aligns each spectrum to a reference spectrum. So, that the baseline and the amplification effects are at the same level in every spectrum.

(Refer Slide Time: 25:38)

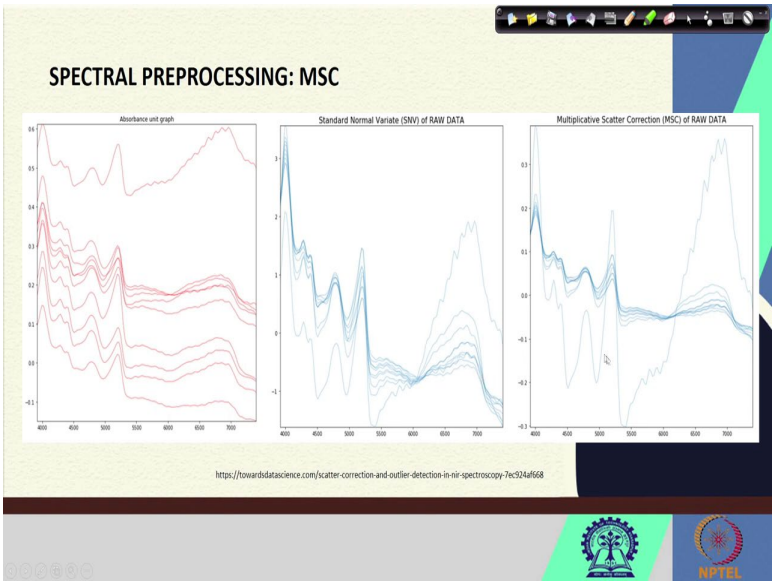


So, what is the way of doing this multiplicative scatter correction because this reference spectrum is unknown. So, the mean spectrum of a given spectral library denoted, so, generally we consider a mean spectrum which is denoted by this xref. So, these xref represents the mean scattering and offset and then each spectrum Xi then fitted to the reference spectrum using the least square method.

We know the least square method that is Xi i plus bi xref plus eta, where these a is offset, bi is the slope and then e is the residual. Now, once we calculate these then we calculate the MSC by using this formula where ai and bi. So, ai is offset bi is the slope. So, using this is the formula for multiplicative scatter correction. So, particle size and path length effect should vary randomly from sample to sample and therefore, the average should reasonably reduce the effects.

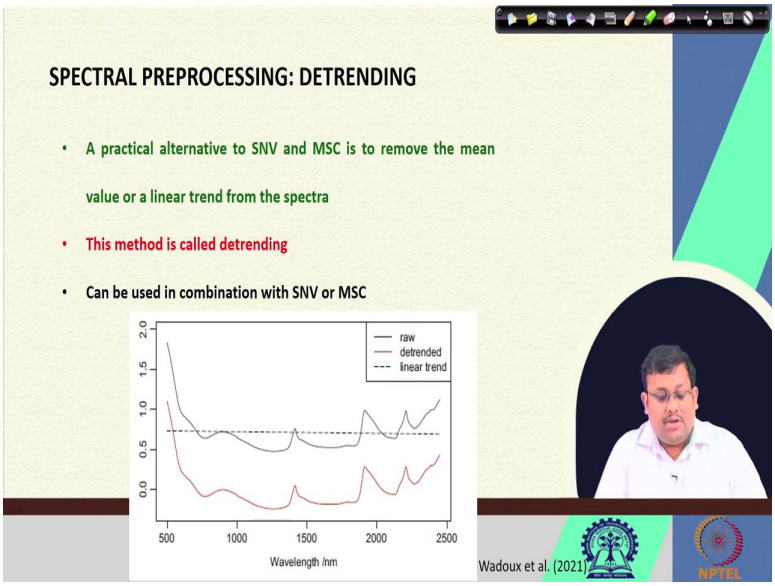
So, this is the fundamental assumption behind this multiplicative scatter correction. So, in case of multiplicative scatter correction, we have seen the basic assumption is the particle size and path length effects should vary randomly from sample to sample and as a result the average should reasonably reduce this effect. So, based on this assumption, this multiplicative scatter correction is generally done. And these two multiplicative scatter correction and SNV are helpful for removing the light scattering effect from the spectrum.

(Refer Slide Time: 27:20)



Now, here you can see some example of these, this is an example of raw absorbance spectra and this is the SNV spectra of these data and also these the MSC spectra of this data. So, you can see the difference from these raw spectra and then MSC and SNV and this MSC spectra.

(Refer Slide Time: 27:43)

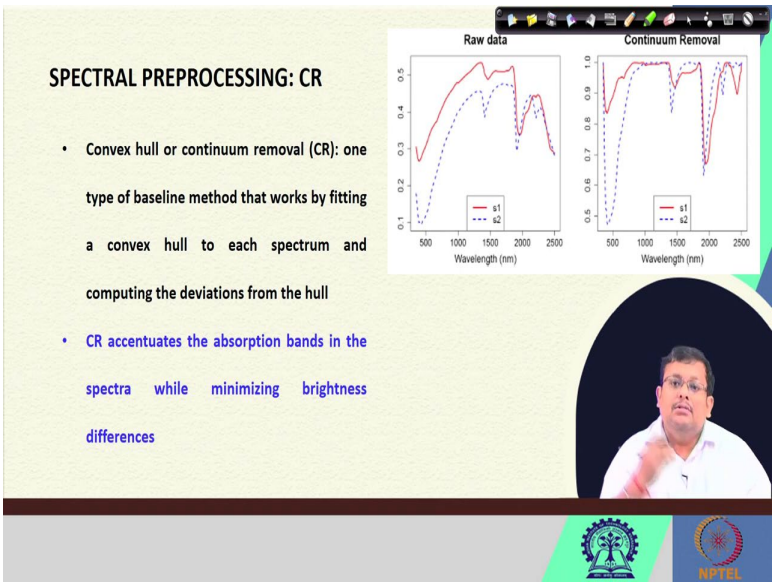


Another method which we generally use is called the detrending method. So, it is a practical alternative to SNV and MSC. And also, generally what we do in case of detrending? So, we remove the mean value or a linear trend from the spectra. So, that is why it is called detrending. And it can be used in combination with SNV and MSC also.

So, generally you will see that some of the, in some literature they have used SNVDT that means standard normal variate followed by detrending. So, sometimes these two spectral pre-

processing are combined together for the ultimate, for the whole spectral pre-processing of the data. So, you can see how these raw data and this detrend data looks like. And this is a linear trend the dashed line is showing the linear trend. So, this detrending is another important spectral pre-processing which is often used in combination of SNV and MSC.

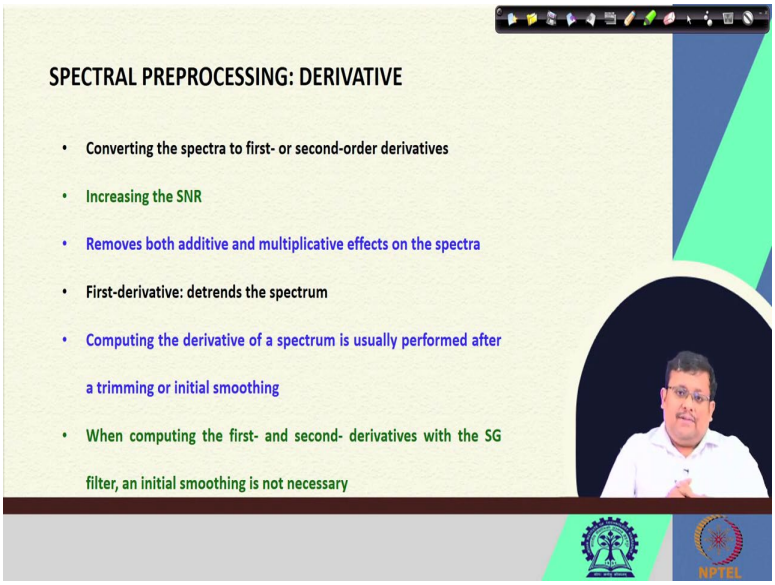
(Refer Slide Time: 28:57)



So, here you can see another spectral pre-processing that is called the continuum removal. So, this convex hull or continuum removal is one type of baseline method that works by fitting a convex hull to each spectrum and computing the deviation from the hull. So, here you can see these raw reflectance data from the soil. So, if we fix a convex hull and then computing the deviation from the hull, then we can get this type of spectral data.

So, this is called the continuum remove spectrum. So, why we go for the continuum remove spectra? Because this continuum removal accentuates the absorption bands in the spectrum while minimizing the brightness differences. So, this is how the continuum removal is also considered as one of the important spectral pre-processing and has been used in different application.

(Refer Slide Time: 30:03)



So, guys, let us wrap up our lecture here. The derivative spectral pre-processing we will be discussing in our next lecture. So, I hope that you have already got some useful information for spectral pre-processing, you have seen what is field of view, what is a large field of view, what is small field of view, how to calculate the field of view and also you have seen what are the important spectral pre-processing we have started with reflectance to absorbance conversion, then spectral trimming, then we have seen continuum removal, then SNV, then detrending, then we have seen multiplicative scatter correction.

So, I hope that you have gathered some useful information and we will start from here, we will discuss the derivative spectral pre-processing in our next lecture, which is one of the most important spectral pre-processing methods. So, we will discuss the derivative spectral pre-processing, what are the advantages of derivative spectral pre-processing in our next lecture. And we will also see some examples of the soil spectroscopic application for soil specific or soil sand studies.

(Refer Slide Time: 31:31)

So, some of the -- These are the references which I have used; Barnes, which we have given, Barnes et al. who have given the standard normal variate and de-trending. And also, the Reyna et al, you can see that some examples are given. And again Wadoux et al. you should follow this book.

I would highly recommend you should follow this book where you can have some R packages and the already these datasets are there which you can plot and you can have more confidence of dealing with the spectral pre-processing and you can see what things are changing, what aspects are changing when you transfer one soil spectra with a specific type of spectral pre-processing.

And you will see by yourself and you will gain enough confidence by doing this spectral pre-processing. And if you want to know more about the spectral pre-processing and how to do, you should follow this book. So, it is highly recommended that you should consider this book for gaining more information regarding the different types of spectral pre-processing.

So, thank you guys and let us meet in our next lecture where we will see the derivative spectroscopy. And also, we will be discussing about some of the important application, soil application of diffuse reflectance spectroscopy. Thank you very much.