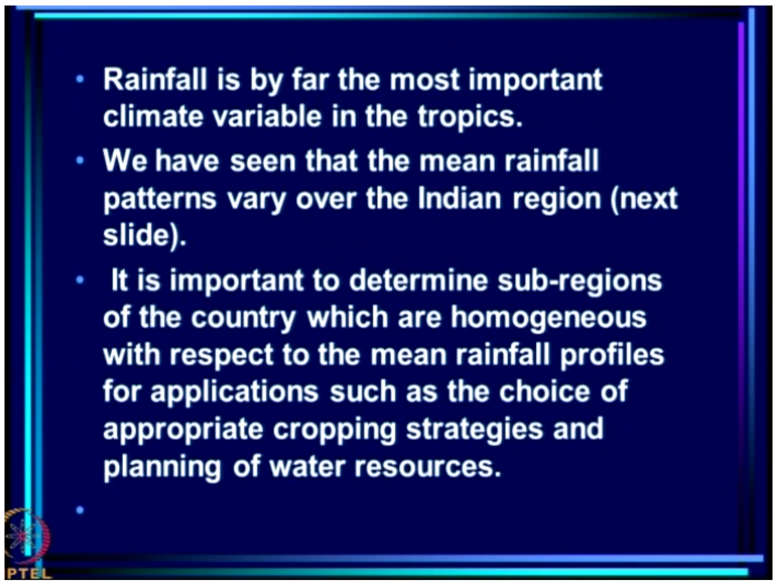
The Monsoon and Its Variability
Prof. Sulochana Gadgil
Center for Atmospheric & Oceanic Sciences
Indian Institute of Science – Bangalore

Lecture - 19
Climatic Clusters of the Indian Region

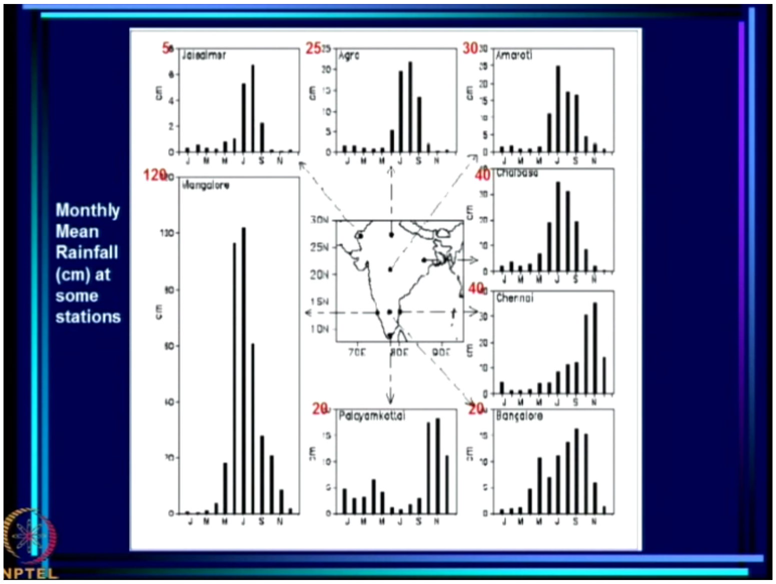
So, today I will talk about the climatic clusters of the Indian region.

(Refer Slide Time: 00:18)



Now, although I said climate actually we will talk primarily about rainfall because rainfall is by far the most important climate variable in the tropics. Now we have seen that the mean rainfall patterns vary over the Indian region.

(Refer Slide Time: 00:37)

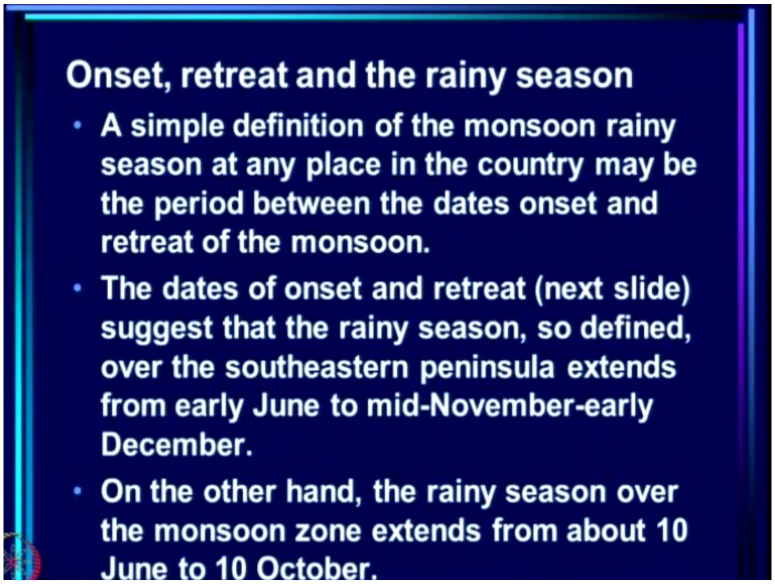


This we have seen before that there is considerable variation in the mean rainfall pattern. So, this is the typical west coast pattern. Here over Mangalore and then this is Chennai or what used to be called Madras and these are very different and then you have the monsoon zones with characteristic high rainfall during June, July. Now, why do we have to worry about climatic clusters?

See, it is important to determine sub-regions of the country which are homogenous with respect to the mean rainfall profiles for applications such as the choice of appropriate cropping strategies and planning of water resources. So, in fact this problem is given to us by a scientist from the International Crop Research Institute for semiarid tropics saying that if you want to think about the optimum cropping patterns the first thing we think off is what is the mean rainfall pattern like so and it is the same or similar in place A from place B and it would be so if they belong to the same homogenous region.

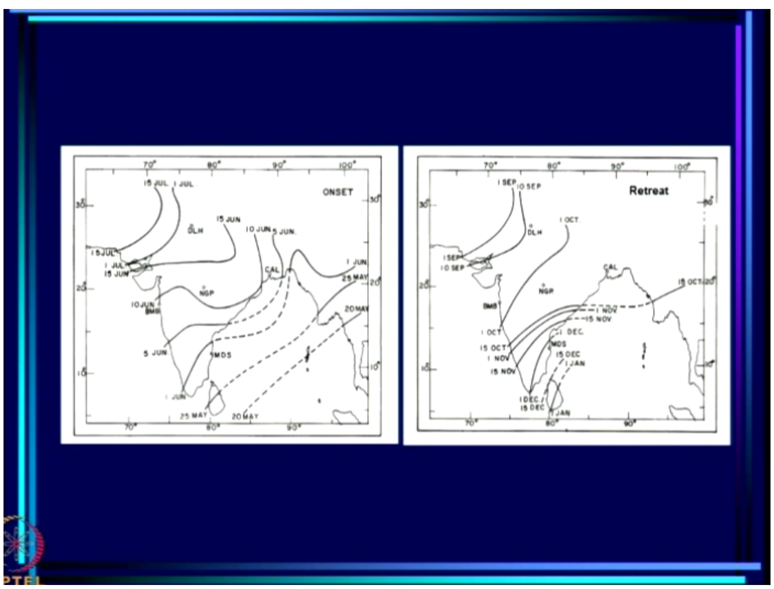
So this is why we try and determined the homogenous sub-regions of the Indian region which are homogenous relative to profiles rainfall patterns like you have seen here and these remember are patterns of the mean rainfall. This is mean monthly rainfall.

(Refer Slide Time: 02:04)



Now we have seen that the rainy season varies from place to place and a simple way to define the rainy seasons or the monsoon rainy season at any place in the country may be the period between dates of onset and retreat of the monsoon. So, the dates of onset and retreat

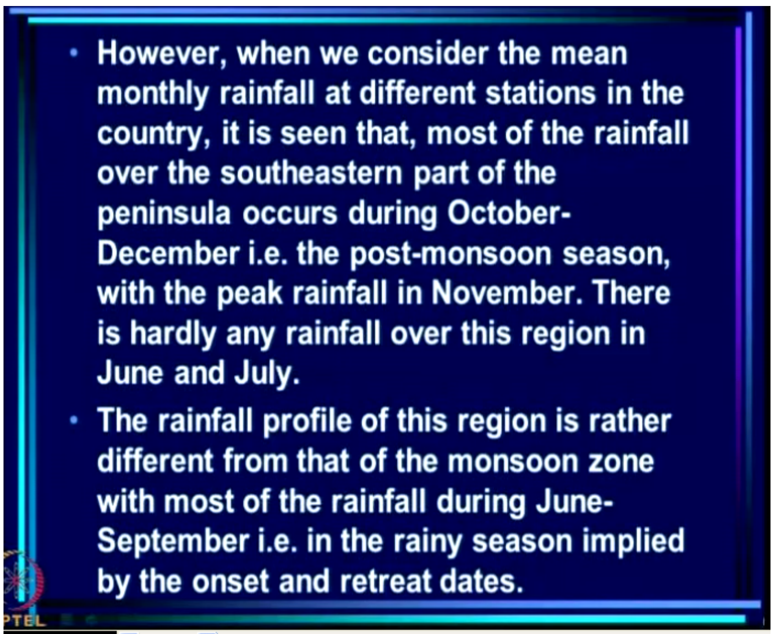
(Refer Slide Time: 02:26)



is something we have seen before and so these are the dates of onset and what you see here is that the onset varies date varies from 1st June here to 15th July and the dates of retreat vary from 1st September to 15th of December here. 15th of December is outside India to about early December in this part of Tamil Nadu coast. So the dates of onset and retreat suggests that the rainy season so defined over the south eastern peninsula extends from early June to mid-November to early December.

So, this is what we have seen here that the rainy season between onset and retreat really is very long from early June to almost mid-November or early December. On the other hand, the rainy season over the monsoon zone extends from about 10th of June to 10th of October. You can see here that by 10th of June the onset has occurred here after 10th June the onset occurs in this reason so the rainy season depending on where you are in the monsoon zone, extension of 10th June to October or so.

(Refer Slide Time: 03:40)

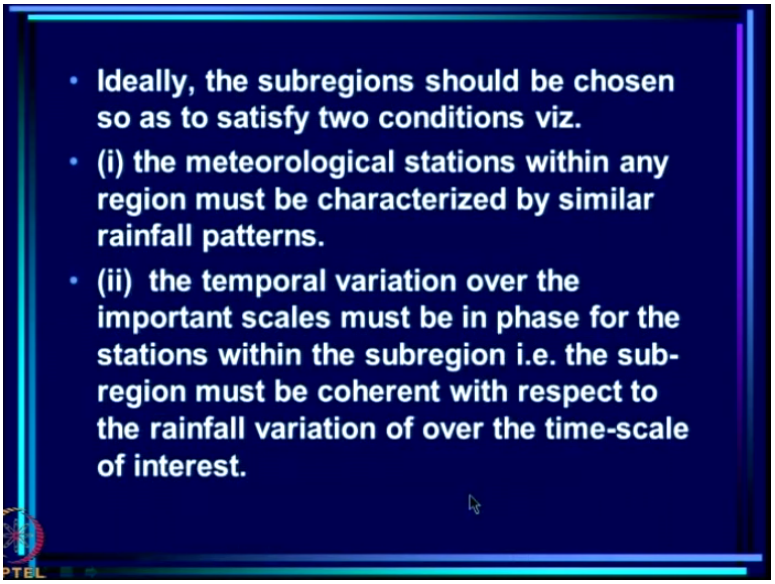


So the rainy season as defined by onset and retreat dates are rather different for these 2 regions. However, when we consider the mean monthly rainfall at different stations in the country it is seen that most of the rainfall over the south eastern part of the peninsula occurs during October to December that is the post-monsoon season with the peak rainfall in November. There is hardly any rainfall over this region in June and July.

So the rainfall profile of this region is rather different from that in the monsoon zone. So we have seen this already that if you look at rainfall pattern of place like Chennai then most of the rain occurs in October, November although IMD has declared the rainfall begins around June there is hardly any rain in June. Most of the rain occurs here in October, November and yet if you went by the definitions of onset and retreat we would have a rainy season which extends so many months here.

So the rainfall profile of this region is rather different from that of the monsoon zone which we have seen.

(Refer Slide Time: 04:52)

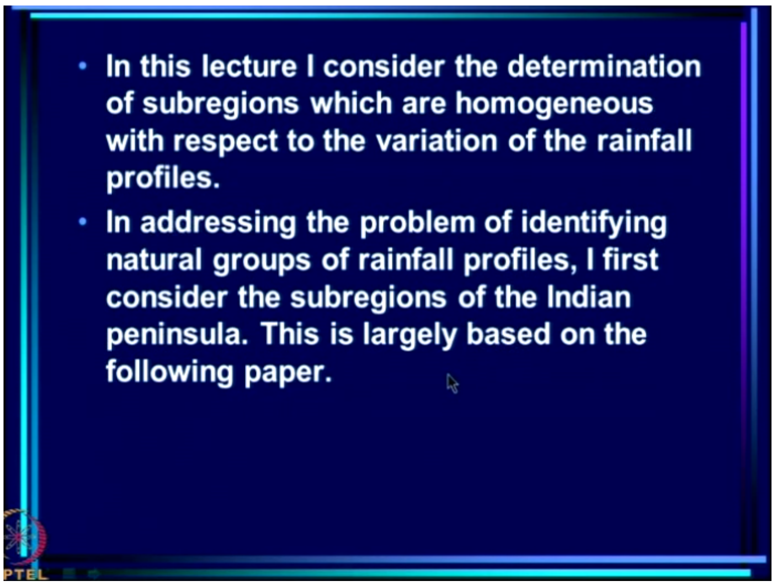


So ideally so this suggests that we have to take a relook at these the definition of rainy season also and I will come back to this. So now we are trying to look at sub-regions which are homogenous relative to this rainfall pattern this variation of mean rainfall over the region. Now, ideally the sub-regions should be chosen so as to satisfy 2 conditions, namely that the meteorological stations within any region must be characterized by similar rainfall patterns and this is what I am going to address in this lecture.

Secondly, the temporal variation over the important scales must be in phase for the stations within the sub-region that is the sub-region must be coherent with respect to rainfall variation. So there are 2 different aspects here we are saying the mean rainfall pattern has to be similar, but here we are saying that not only the mean has to be similar, but the variation over the important time-scale say, for example is season must be in phase for the stations within the region that is to say.

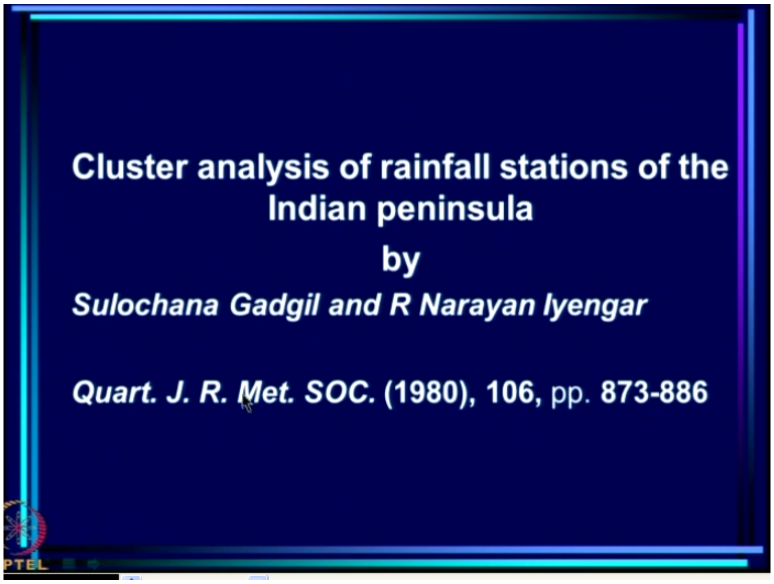
If 1 station gets excess rainfall during the summer monsoon the other stations should not have deficit rainfall that is to say the variation should be coherent and in fact I am not likely to have time in this lecture series to talk about the coherent rainfall zones if I do find that I will include a lecture on coherent rainfall zones but as far as this lecture is concerned we will restrict ourselves to climatic clusters or determining sub-regions which are characterized by similar rainfall patterns.

(Refer Slide Time: 06:39)



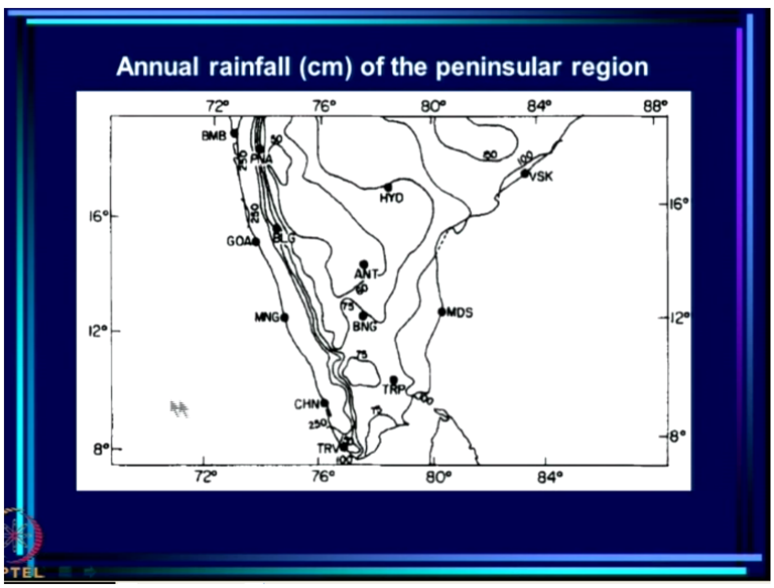
So, in this lecture I consider the determination of sub-regions which are homogenous with respect to the variation of the rainfall profiles. In addressing the problem of identifying natural groups of rainfall profiles I first consider the sub-regions of the Indian peninsula.

(Refer Slide Time: 06:56)



And this is largely based on this paper, cluster analysis of rainfall stations of the Indian peninsula, which was published by in quarterly journal of Royal Met Society in 1980.

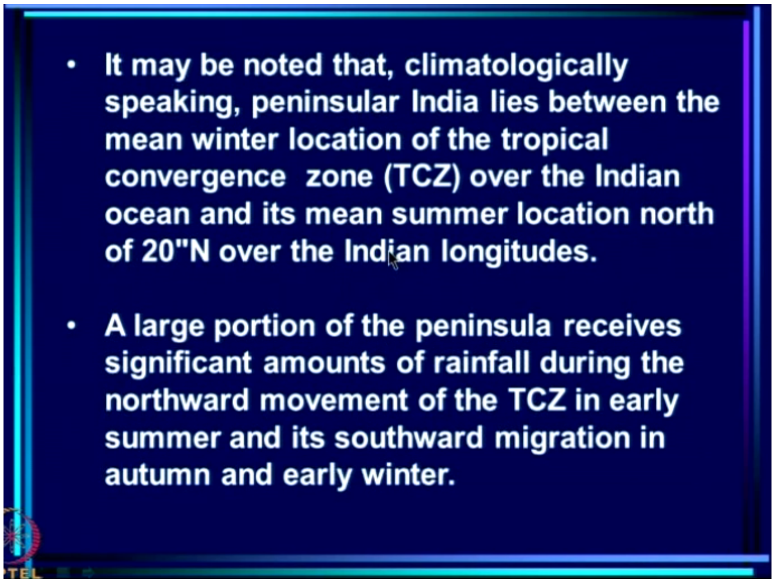
(Refer Slide Time: 07:07)



So, we are now considering the peninsular region only and these are the stations that we have considered in the peninsula. Now, this just gives you a view of how the annual rainfall varies in the peninsular region. You see very heavy rainfall larger than 250 cm over the west coast region. Now, the rainfall decreases as you come this side of the Ghats and in fact it is very low over this region. This is Anantapur about which we will talk later. It is a semi arid part.

This is only 50 and 50 cm around here and it is < 60 cm over this whole region here. Then it increases somewhat. It is about 75 in Bangalore and again in Madras and so on. It is over 100 cm. So, we have somewhat higher rainfall here relative to the center of the peninsular region.

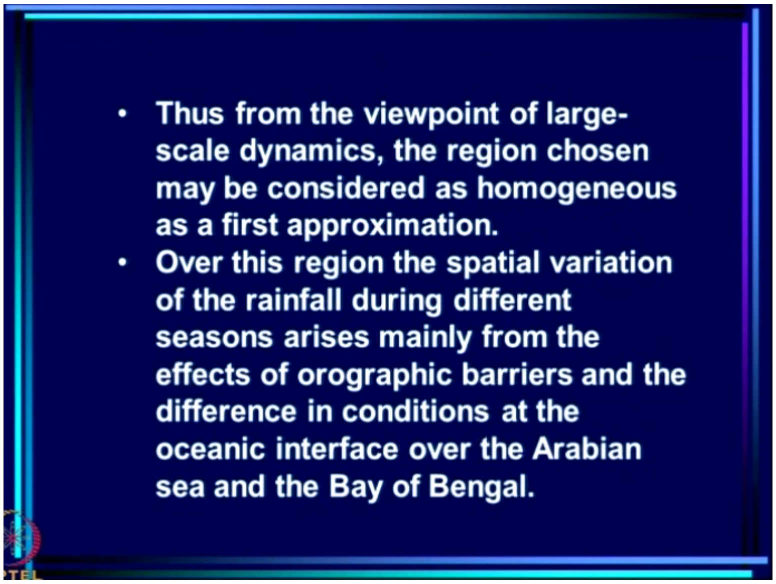
(Refer Slide Time: 08:08)



Now we have to also remember another thing if we look at the peninsular region, we know very well that this peninsular India lies between the mean winter location of the TCZ over the equatorial Indian Ocean and its mean summer location north of 20 degrees North over the Indian longitudes. So, this peninsular region is between the mean winter location over the equatorial of the equatorial TCZ and the mean summer location over the monsoon zone.

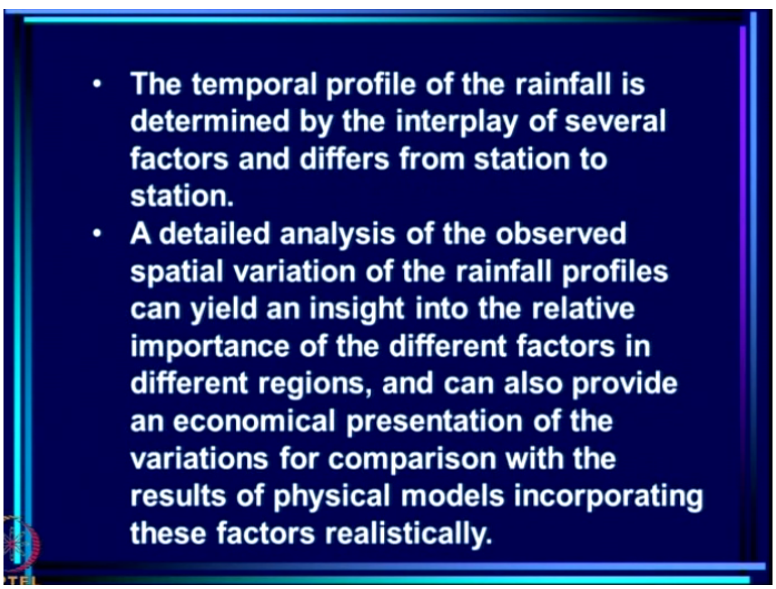
Now large portion of the peninsular receives significant amount of rainfall during the northward movement of the TCZ in early summer. This is to say in the onset phase and its southward migration in autumn and early winter which is the retreat phase. So the peninsula does receive a lot of rain during this transition phases.

(Refer Slide Time: 09:05)



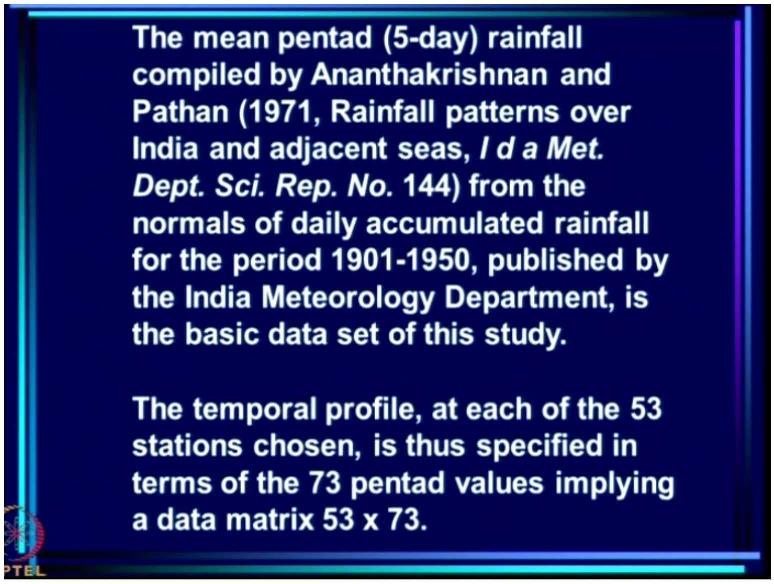
So from the point of view of large-scale dynamics, the region chosen may be considered as homogenous as a first approximation. Over this region, the spatial variation of rainfall during different seasons arises mainly from effects of orographic barriers and the difference in conditions at the oceanic interface over the Arabian sea and the Bay of Bengal.

(Refer Slide Time: 09:30)



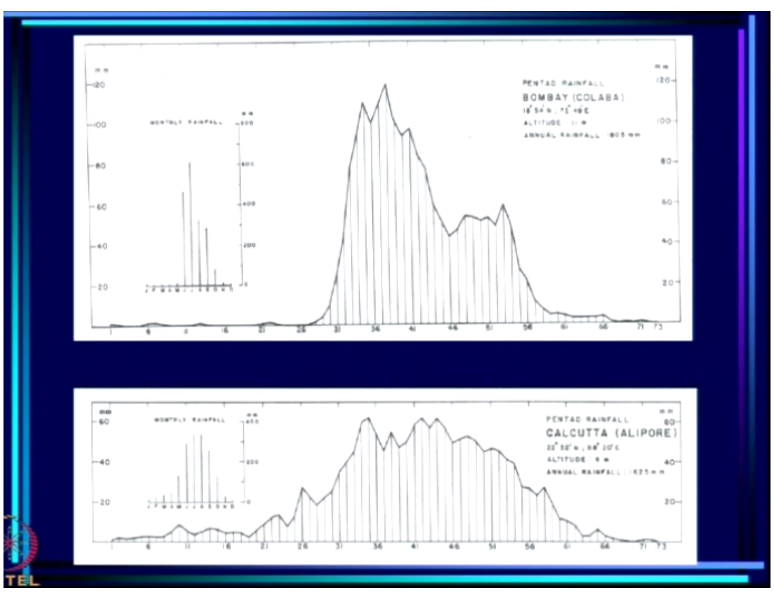
The temporal profile of the rainfall is determined by the interplay of several factors and differs from station to station. A detailed analysis of the observed spatial variation of the rainfall profiles can yield insight into the relative importance of the different factors in different regions and can also provide an economical presentation of the variations for comparison with the results of physical models incorporating these factors realistically. So in many ways this kind of identification is useful.

(Refer Slide Time: 10:01)



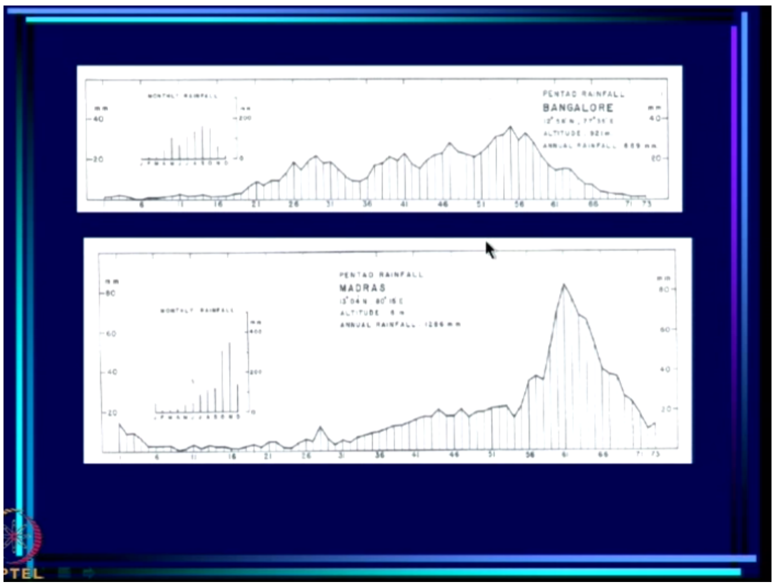
Now the mean pentad which is 5-day rainfall complied by Ananthakrishnan and Pathan in 1971 is used as the basic data and this was compiled from the normals of daily accumulated rainfall from the period 1901 to 1950 published by India Met Department. Now the temporal profile at each of the 53 stations chosen is thus defined or specified in terms of 73 pentad values. There are 73 pentads in a year.

(Refer Slide Time: 10:33)



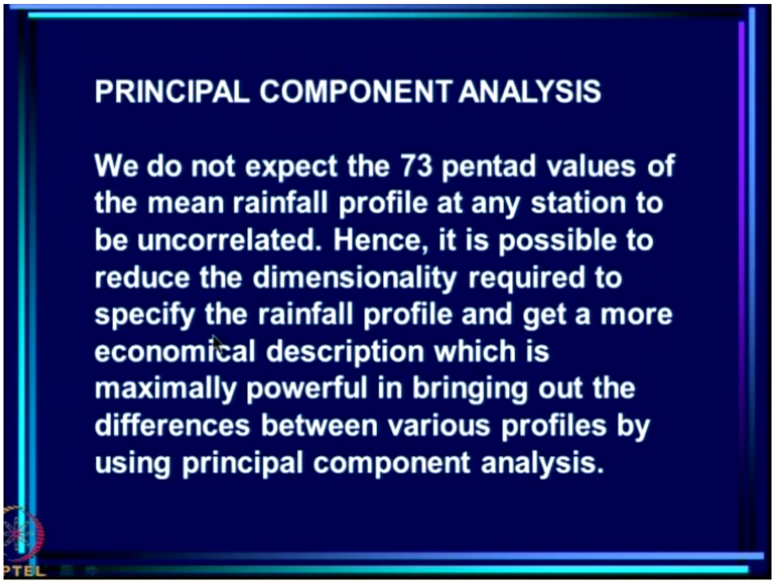
And just to give you a feel for the kind of the data we have this is the kind of data we have. This is the monthly rainfall shown here and this is for Bombay which is now known as Mumbai of course and you see a very sharp onset in the pentad data and this is the pattern. So, this is the pattern for Bombay. This is the pattern for Calcutta. You see much more sprawling pattern and much less sharp onset as far as the mean pattern is concerned

(Refer Slide Time: 11:03)



This is Bangalore which is sprawling even more and this is Chennai which has a very sharp relatively sharp onset towards the end of September and very sharp peak around October and November. So this is Chennai. So these are the basic data that have been analyzed to determine the clusters.

(Refer Slide Time: 11:28)

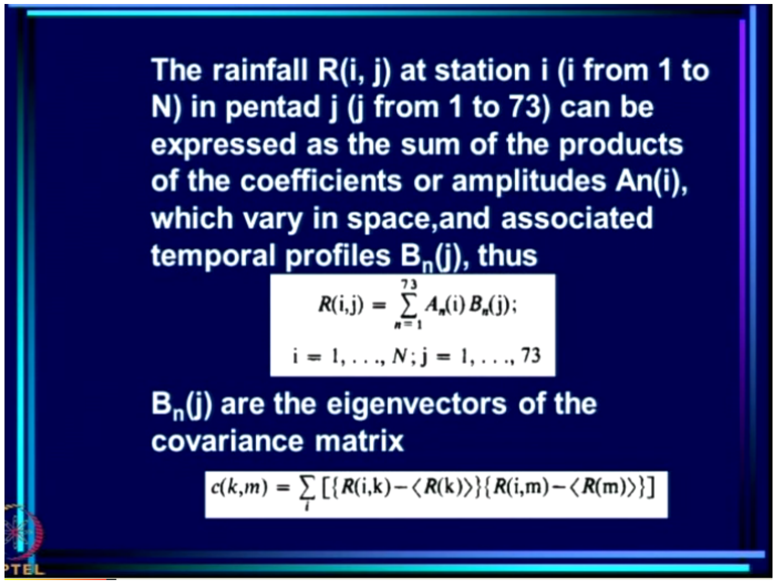


These are 5 data. Now as you can see, see we do not expect that the rainfall in any pentad will be uncorrelated with the rainfall in the previous pentad not at all because the season is characterized by several pentads in which for example it increases here and several pentads in which it is more less steady with a gentle increase here.

So, we expect in general that rainfall in any pentad will be correlated with the rainfall on the previous pentad, next pentad, and so on. So these 73 values that represent the rainfall pattern are by no means independent. So we do not expect the 73 pentad values of the mean rainfall profile at any station to be uncorrelated. Hence it is possible to reduce the dimensionality required to specify the rainfall profile and get a more economical present description which is maximally powerful in bringing out the differences between the various profiles by using principal component analysis.

Remember when we try and determined homogenous groups we want a tool which will bring out the differences very effectively so that we can make sure that what we have identified a similar profile are indeed similar.

(Refer Slide Time: 12:46)

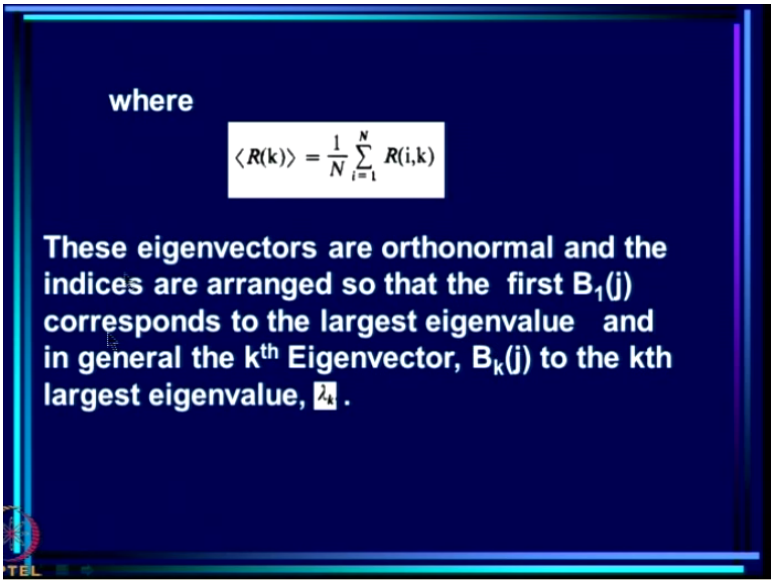


Now this is the principal component analysis while I do not plan to go in detail on this. There are many references available. Let me just explain what the method involves. So, the rainfall at a station i and for i = 1 to n the number of station and in pentad j, j going from 1 to 73 can be expressed as the sum of products. So what we are doing now is instead of taking 73 pentad values, we are taking in fact modes which involve weightages to all the 73 values.

And those modes are multiplied by amplitude so the rainfall is represented as a sum of amplitudes An(i) which as specific to this station and temporal profiles Bn(j). These temporal profiles are determined from the entire data set. These are profiles determine so that they explain maximum variance of the rainfall from station to station. So R(i, j) then is the sum over all these j. n = 1 to 73 over all the pentads of An(i) and Bn(j). Bn(j) are the patterns temporal patterns that we have derived and these are the coefficients.

So Bn(j) are in fact the eigenvectors of the covariance matrix and you know that covariance matrix c(k, m) is simply a product of the deviations. So, R(i, k) - the mean at k and R(i. m) - the mean for m, m being the pentad.

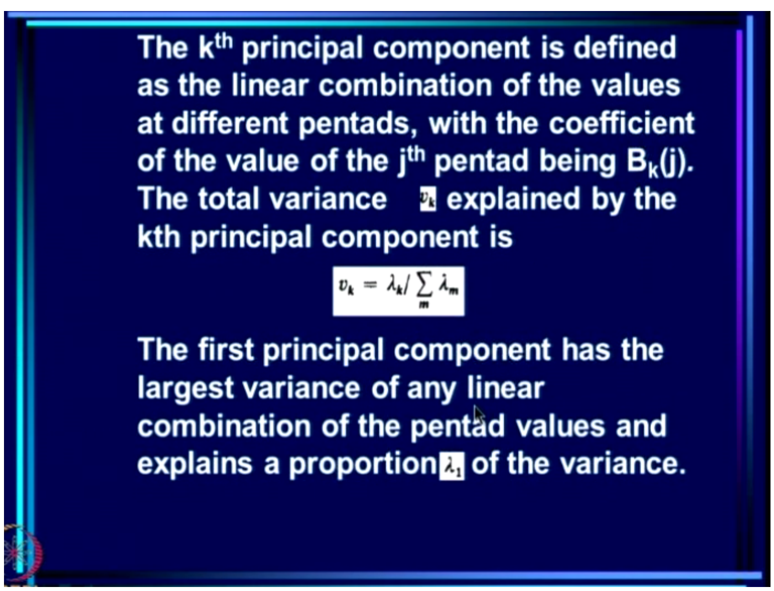
(Refer Slide Time: 14:35)



Here summed over all the pentads. So N equal to so where R(k) is just the mean which is i equal to 1 to N R(i , k). So this is the mean rainfall for that pentad k and taking this mean and looking at the departures from the mean we can get the covariance matrix which we see here. It really does not matter. The point is that these eigenvectors are orthonormal okay.

So, they are perpendicular to one another and the indices are arranged so that the first which is B1(j) corresponds to the largest eigenvalue and in general the kth Eigenvector, Bk(j) to the kth largest eigenvalue lambda k okay. So these are organized such that

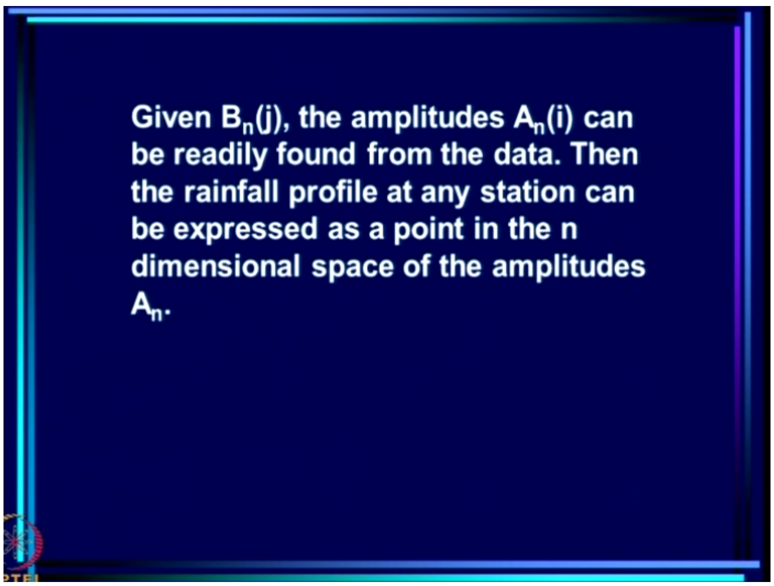
(Refer Slide Time: 15:27)



B1(j) corresponds to the largest eigenvalue Now the kth principal component is defined as the linear combination of the values at different pentads with the coefficient of the value of the jth pentad being Bk(j). So this gives you a temporal pattern and the total variance is vk explained by the kth principal component is vk = lambda k which is the eigenvalue of the kth one divided by the sum of all the lambdas.

The first principal component has the largest variance of any linear combination of the pentad values and explains a proportion lambda 1 of the variance.

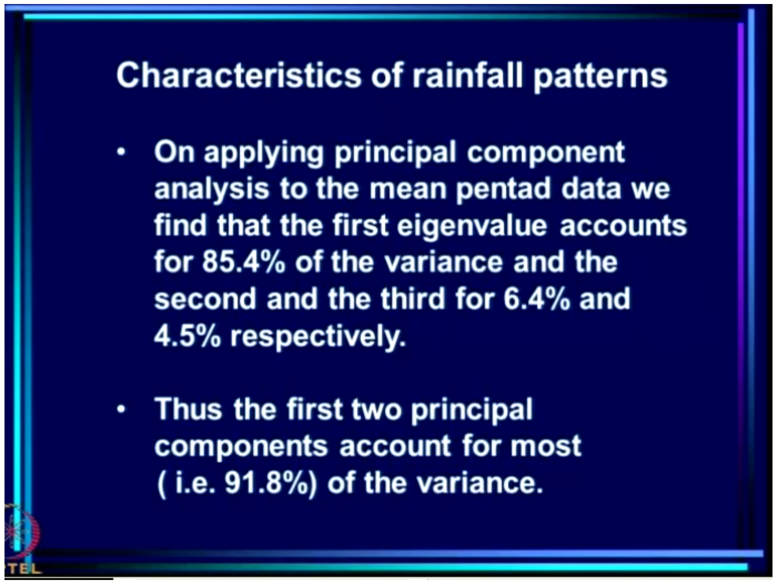
(Refer Slide Time: 16:05)



So this method is extremely effective in economically representing the variation from station to station that we have seen. Given Bn(j) the amplitudes An(i) can be readily found out from the data. Then the rainfall profile at any station can be expressed as a point in the n dimensional space of amplitudes.

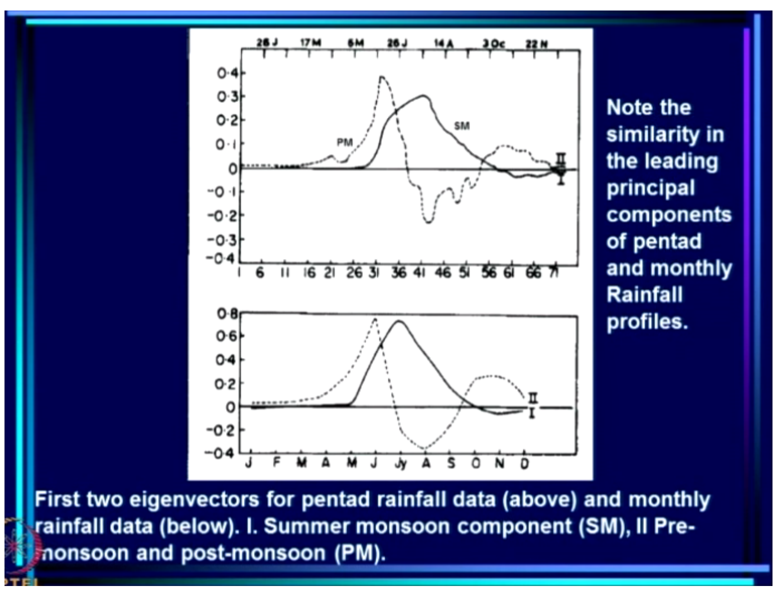
So instead of talking of rainfall in a specific pentad we now talk of what is the coefficient of a certain rainfall pattern and in terms of these coefficients A(i) actually the data can be exactly represented in a n dimensional space okay. So where you have instead of 73 pentad values you have 73 patterns and corresponding to that 73 amplitudes, but the whole point of doing this exercise is you do not have to go to 73 dimensions at all.

(Refer Slide Time: 16:59)



In fact, first 2 or 3 components generally explain most of the variance. So we can actually analyze the variance in a 2 dimensional, 3 dimensional space. So on applying principal component analysis to mean pentad date we find that the first eigenvalue accounts for 85.4% of the variance, second and third 6.4 and 4.5 respectively. So the first 2 principal components account for most of the variance almost 92% of the variance.

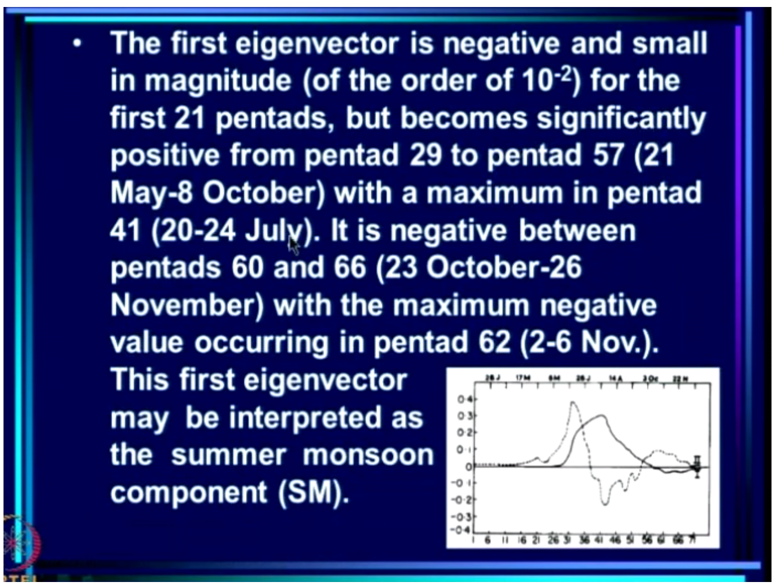
(Refer Slide Time: 17:34)



Since that is the case, now let us see what the components look like and what you see here is the components determine on top from the pentad data and here from monthly data which is also available for all those station. So, the first component is this one which can be called the summer monsoon component because it is maximum in the middle of the year June, July, August. So this is the summer monsoon component.

On the other hand, the second component is what we can call the pre-monsoon and the post-monsoon because this is maximum before the monsoon April, May and also after the retreat of the monsoon from the monsoon zone which is October, November. So this we can call the south summer monsoon and this can be the post-monsoon and in fact whether we take pentad or we take monthly we get very, very similar patterns. So it is not surprising. So the first 2 components then correspond to the summer monsoon and the post-monsoon.

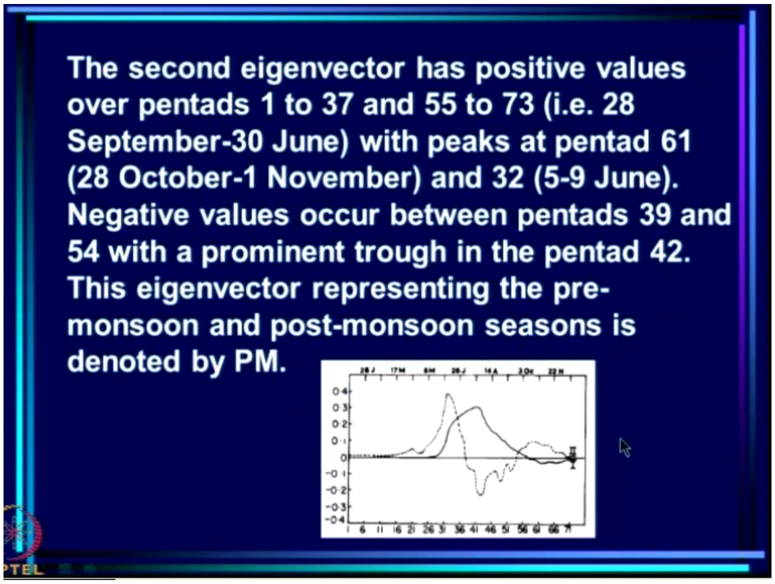
(Refer Slide Time: 18:39)



We look at it. Now notice that the first eigenvector is actually negative, but you can hardly see that it is negative it is almost hugging the axis. It is of the order of 10 to the -2 for the first 21 pentads okay, but becomes significantly positive from 29 to pentad 51 which is from 21st May that is just before the onset to about 8th of October. So this becomes positive in this region here okay and it is negative again between 23rd October and 26th November.

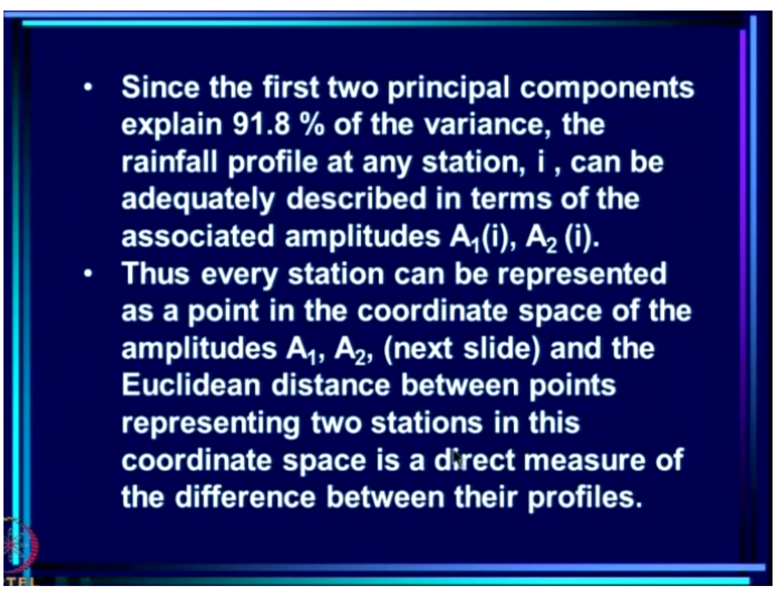
What does negative mean? It means that if both of these have positive coefficients then this will oppose this second component here. This is what negative means. Remember these are just weightages to be given to the different pentads that is what these components correspond to. Now, so the first eigenvector we have interpreted as the summer monsoon one.

(Refer Slide Time: 19:42)



The second eigenvector has positive values from pentad 1 to 37 right up to here okay and 55 to 73 from here onwards again. So essentially from 28th September all the way round till 30th of June this has positive values and negative values occur between pentads 39 and 54. These are the negative values with a prominent negative value in the pentad 42 and they talk of breaks in the monsoon occurring very frequently in this pentad which are reflected in the peninsular rainfall. This eigenvector representing the pre-monsoon and post-monsoon is denoted by PM.

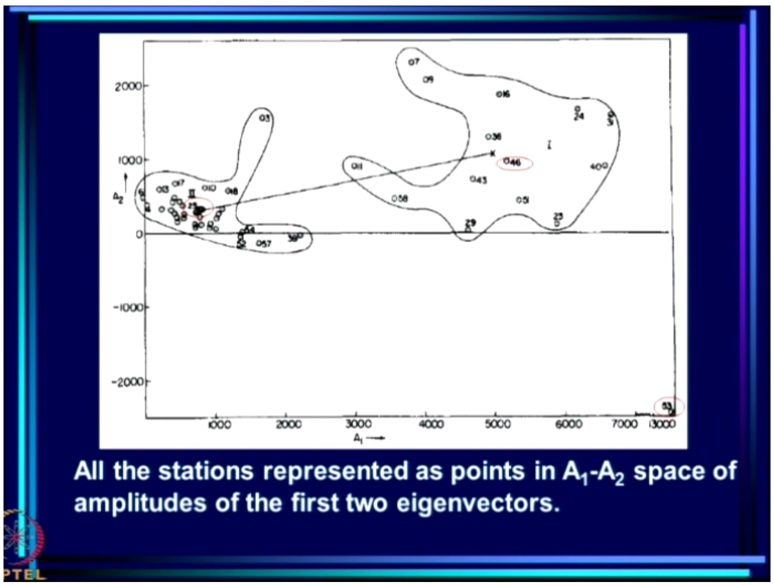
(Refer Slide Time: 20:30)



Since the first 2 components explain 92% of the variance the rainfall profile at any station i, can be adequately described in terms of the associated amplitudes A1(i) and A2(i). So, now we do not worry about the rest of the 71 eigenvectors at all. We say these 2 components have captured most of the variance. So we only look at amplitudes of these 2 components and since 2 components have captured it every station, the profile at every station can be represented as a point in the plane of A1 and A2.

See this is the big advantage of using principal components that actually with 92% of the variance explain we can represent each profile of each station as a point in a 2 dimensional space. Now this is a very important simplification.

(Refer Slide Time: 21:25)



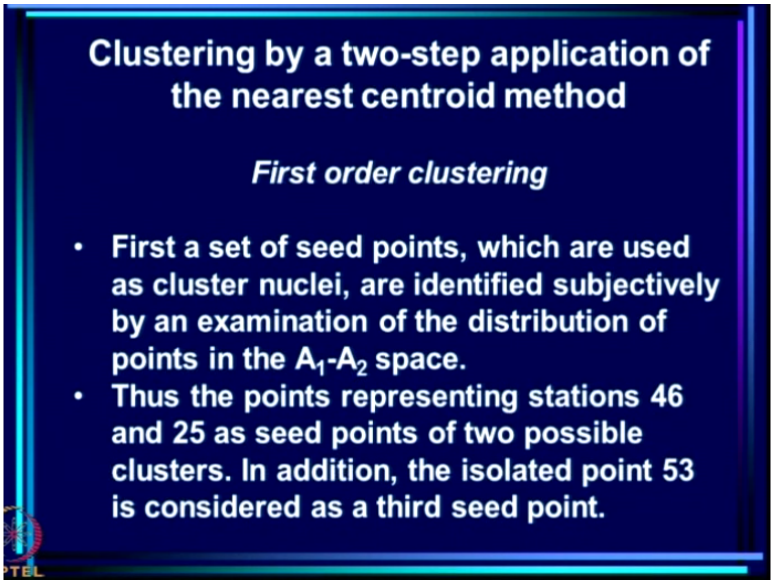
And these are the, this is a one remember this is the summer monsoon component and this is A2 and these are the stations each of then the numbers are given here. They need not interest you. These are the stations that are plotted in the space of A1 and A2. Remember there is also one odd station here, right here with A1 which is very, very large and A2 which is negative.

Now having actually represented the rainfall pattern at each point at each station as a point in the A1 and A2 plane the differences between profiles at 2 stations are just the Euclidean distance between A1 and A2 between the points in the A1 and A2 plane. Representing the 2 stations and is therefore a direct measure of the difference in the profiles. So this is the big advantage now that if you look at distance between any 2 points that is the direct point here.

For example, in a point here the Euclidean distance between the 2 the usual distance x square plus y square, square root of is actually a measure of the difference in the rainfall patterns at these 2 stations. So, given that then it is very easy to determine the clusters because points which are very close together in this face plane in this plane of A1, A2 in fact have very similar profiles and points which are very different from one at very at great distance from one another like a point here and a point here for example they are very different in terms of the rainfall pattern.

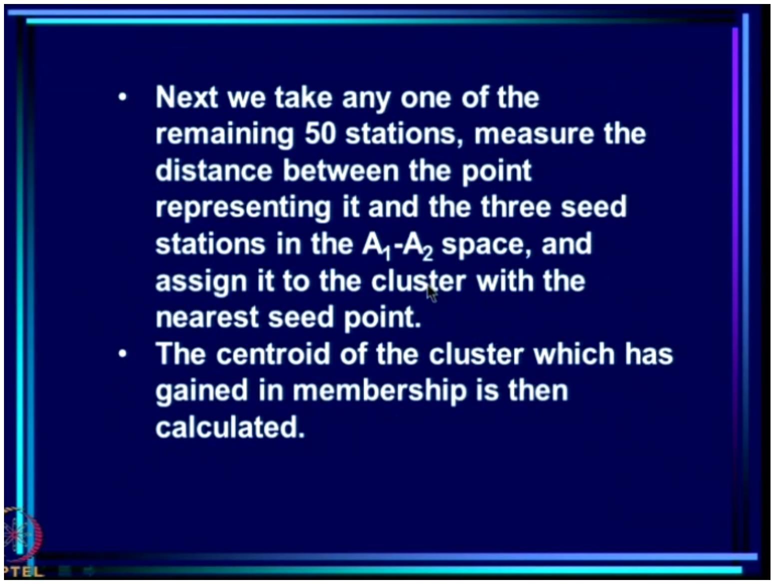
So it is easy then to determine the clusters of the points actually simply by i, but we do not that because that would be subjected to some extent we use a method for determining the clusters.

(Refer Slide Time: 23:22)



The way we cluster it is the following. First, we choose a set of points which are used as cluster nuclei which are identified subjectively by an examination of the pattern. So if you just look at all the points here we see that this is sort of center of this cluster. Here this is the center of this cluster here and that is this isolated point here. So these 3 can be taken as seed points simply by looking at the plot. So we take 46 and 25 as seed points of 2 possible clusters and the isolated point 53 is considered as 1 more.

(Refer Slide Time: 23:98)

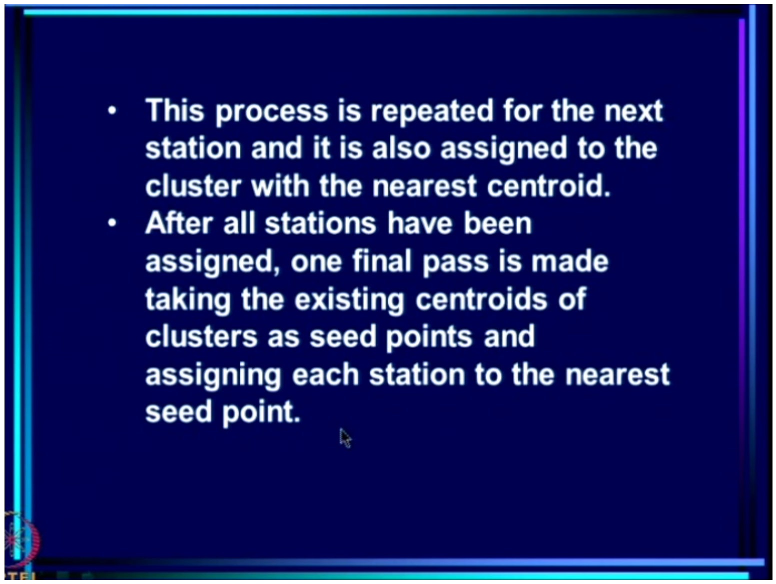


Now, how do we form the clusters having taken the seed points? We take any other station now other than these 3 and measure the distance between the point representing it and the 3 seed stations in the A1 and A2 space and assign it to the cluster with the nearest seed point okay. So, what we do is this is the center here now if the station chosen is here it is obviously closure to this seed point then the others so it will be assigned to this cluster here that is the way we do it.

The centroid of the cluster which has gained in membership is then recalculated. So all the way begins with the seed point here suppose we have now added another station to that then we do not go on looking at measure distances relative to this seed point, but rather we find the centroid of these 2 points which are now members of this cluster and then compare whenever we take the next station perhaps this one the distance from this centroid with the distance from this seed point and 53 this seed point.

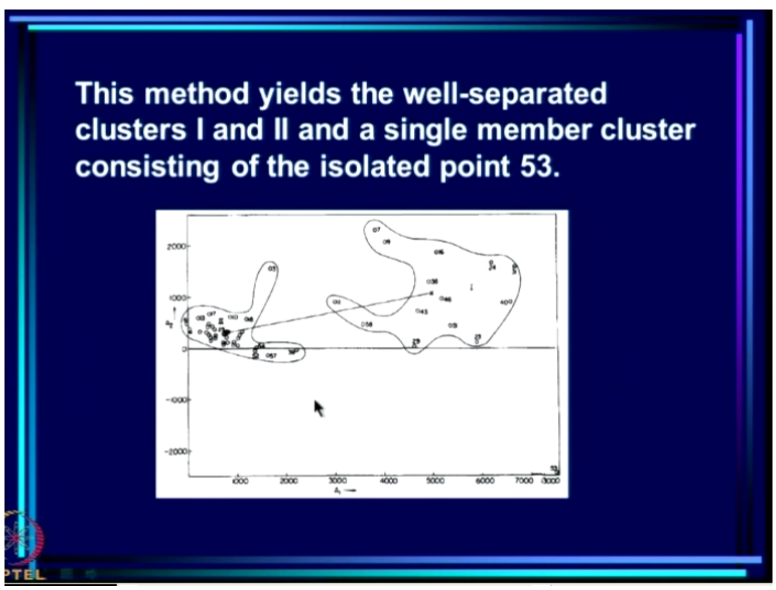
So, the centroid of the cluster which has gained in membership is then calculated.

(Refer Slide Time: 25:22)



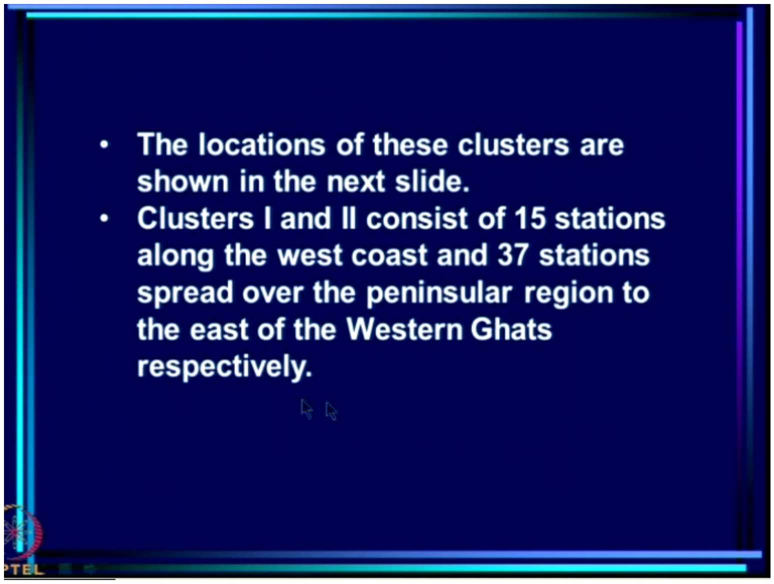
And we go on repeating this process for the next station and it is also assigned to the cluster with the nearest centroid. After all stations have been assigned, one final pass is made taking the existing centroids of clusters as seed points and assigning each station to the nearest seed point.

(Refer Slide Time: 25:42)



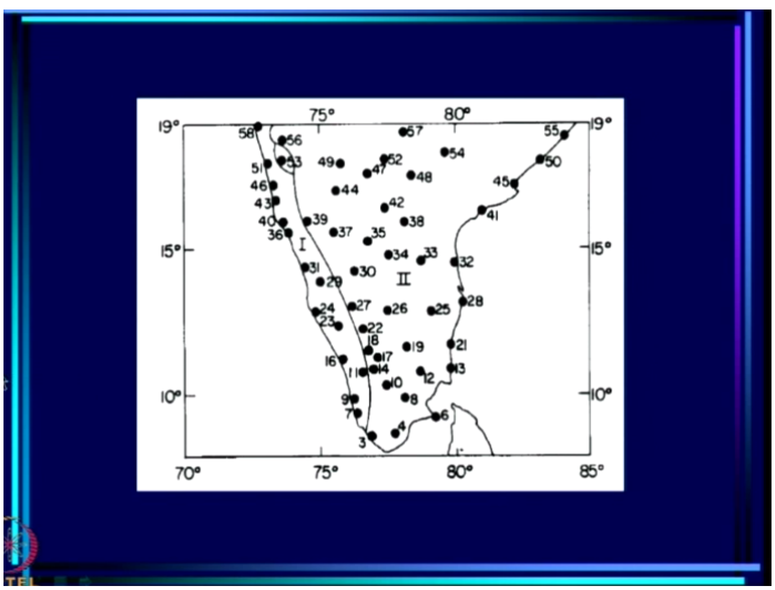
So, this method yields well separated clusters I and I this is 1 cluster here and this is another cluster here and you can see that they are well-separated. There is a big gap between the 2 clusters and this isolated point 53 station 53 here okay.

(Refer Slide Time: 26:00)



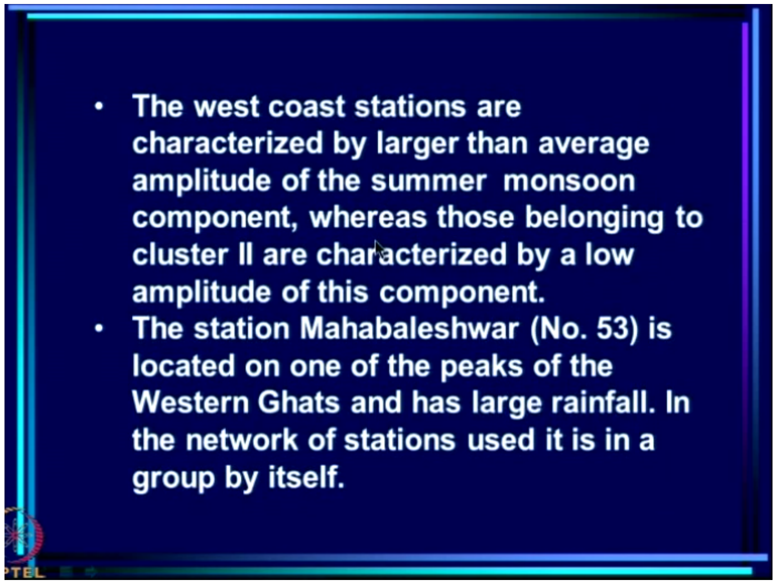
So, the locations of these clusters are shown in the next slide. See so far we have seen station numbers in the A1 and A2 plane these are the locations.

(Refer Slide Time: 26:08)



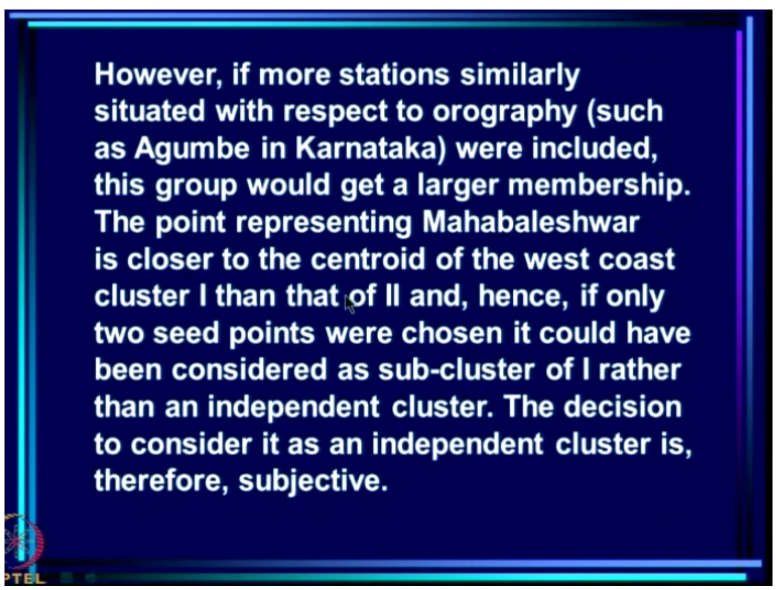
So this, the cluster I then flip act I then this is the cluster II and there is that isolated point 53 is here it is in fact on the crest of the Western Ghats. This station 53 and we do not have other stations which are similar on the crest of Western Ghats in this data set. This is why only isolated one is appearing there. So cluster I comprises west coast primarily west coast and cluster II comprises all the other stations here. The entire south peninsula minus the west coast is cluster II here.

(Refer Slide Time: 26:50)



So, the west coast stations are characterized by larger than average amplitude of the summer monsoon component, whereas those belonging to cluster II are characterized by low amplitude of this component.

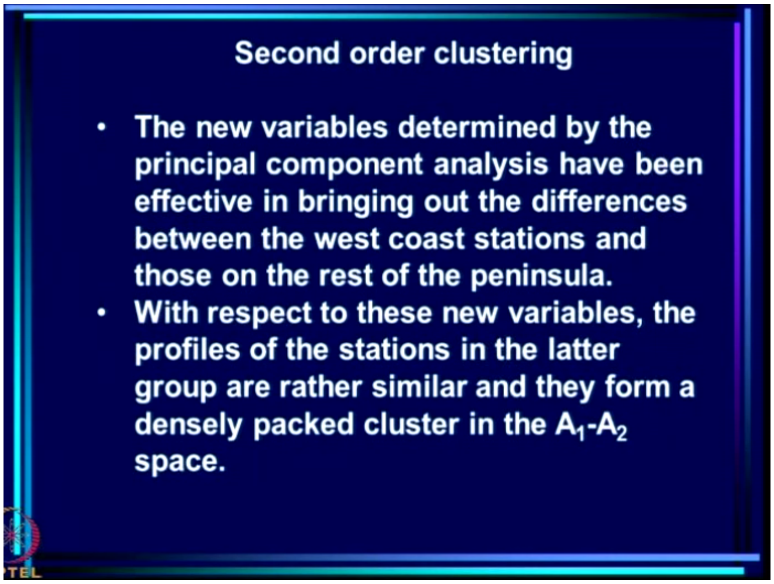
(Refer Slide Time: 27:03)



The station Mahabaleshwar is located. The station Mahabaleshwar is located on the peek one of the peeks of Western Ghats and has large rainfall in the network of stations used it is in a group by itself. However, if more stations similarly situated with respect to orography such as Agumbe in Karnataka were included the group would get a larger membership. The point representing Mahabaleshwar is closer to the centroid of the west coast cluster I then that of II.

And hence if only 2 seed points were chosen it could have been considered as a sub-cluster of I rather than an independent cluster. The decision to consider it independent is therefore subjective.

(Refer Slide Time: 27:47)

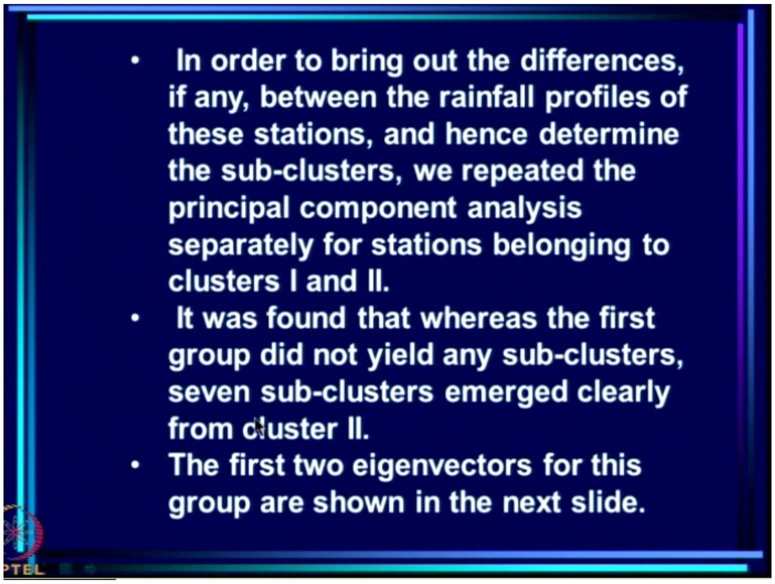


So now we have 2 clusters here cluster I which is the west coast which has you see this is the amplitude of the south monsoon component. This is cluster I so it has a very large magnitude of the summer monsoon component relative to the cluster II which is the entire peninsular region here. Now, entire peninsular region we have seen in fact there are differences in the profiles of the stations in this region.

But they are overwhelmed when we look at the entire thing by the difference between the west coast and the rest and if we look with look at the variation including the west coast station these things appear where close together. So what we now do is to a second order clustering. So we now want to bring out the differences in the patterns rainfall patterns of cluster II which is the peninsula other than the west coast.

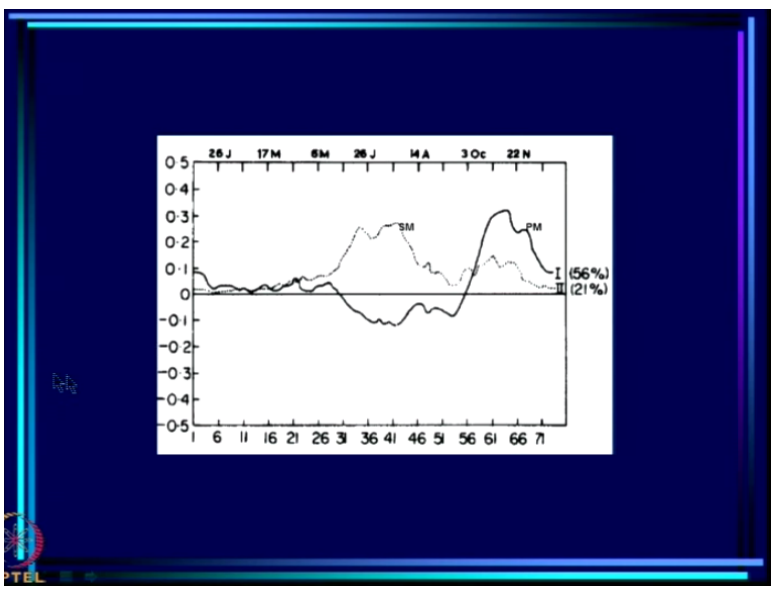
The new variables determined by the principal component analysis have been effective in bringing out the differences between the west coast stations and those on the rest of the peninsula. With respect to these new variables, the profiles of the stations in the latter group are rather similar and they form a densely packed cluster in the A1-A2 space which we have seen.

(Refer Slide Time: 29:14)



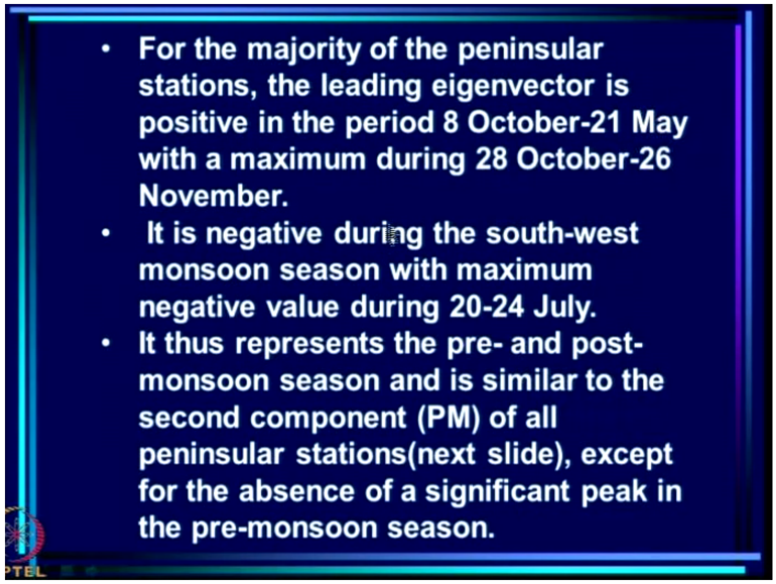
Now in order to bring out the differences if any, between the rainfall profiles of these stations, and hence determine the sub-clusters, we repeated the principal component analysis separately for stations belonging to clusters I and II. Now, it was found that whereas the first group did not yield any sub-clusters.

(Refer Slide Time: 29:45)



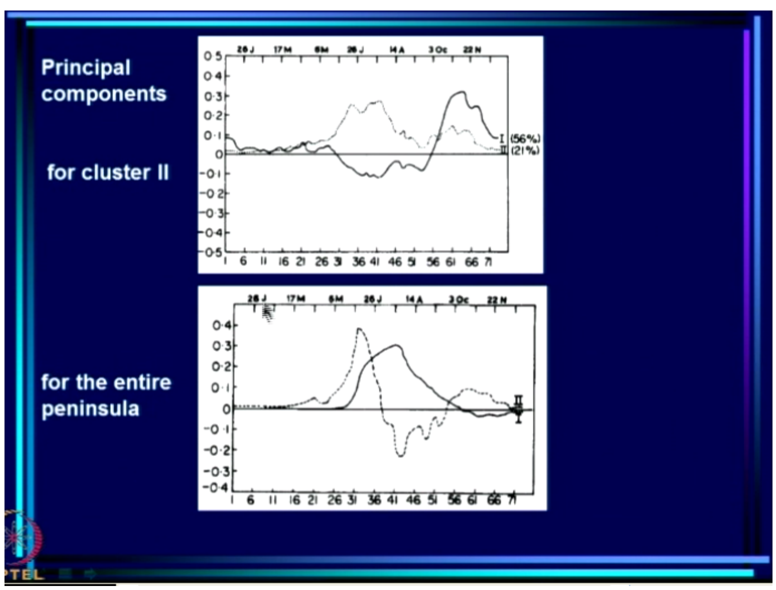
The second group did yield and 7 sub-clusters emerged very clearly from cluster II. So the first 2 eigenvectors of the second sorting which is considering only the stations which are the peninsula stations minus the west coast and what we find is this is the first component and now you see the first component is actually the post-monsoon with peak here and the second component is the summer monsoon. First component explains 56% of the variance, the second component explains 21% of the variance.

(Refer Slide Time: 30:10)



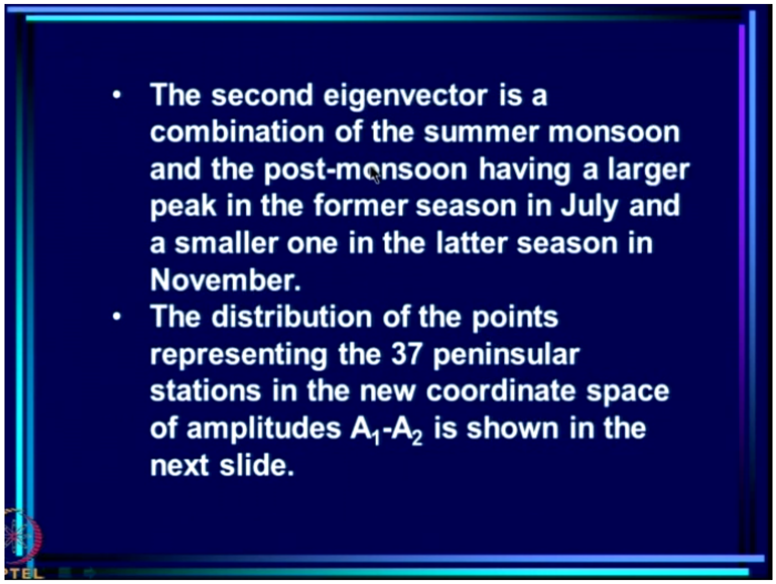
For the majority of the peninsular stations, the leading eigenvector is positive in the period 8th October to 21st May with a maximum during 28th October to 26th November. So this is the maximum here late October to November. It is negative during the south-west monsoon or the summer monsoon season with maximum negative value in here.

(Refer Slide Time: 30:38)



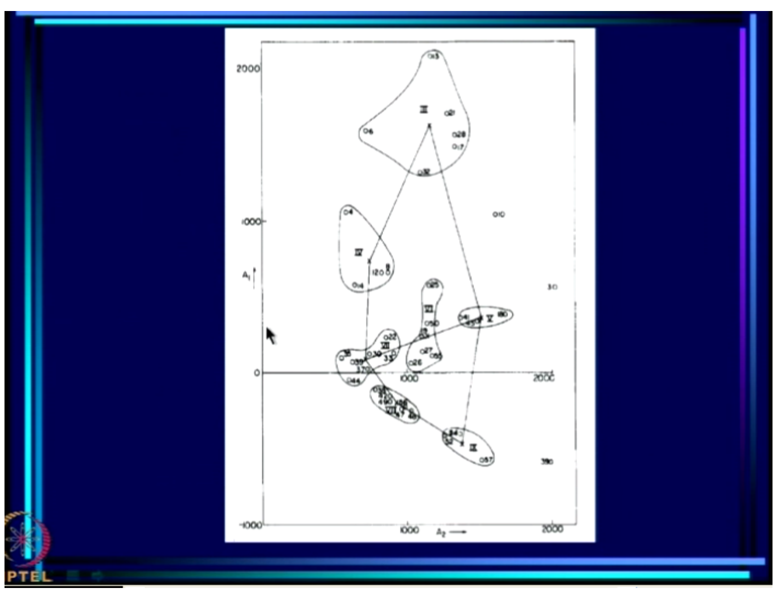
See it is negative during the summer monsoon with maximum negative value here okay.

(Refer Slide Time: 30:46)



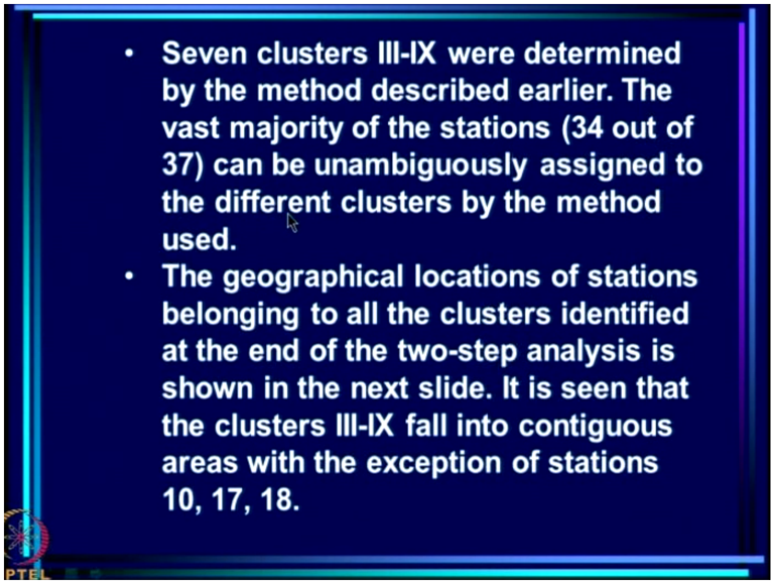
The second eigenvector is a combination of the summer monsoon and the post-monsoon having a peak in the former season in July and a smaller one in the latter season in November. See we see this. So there is one peak here in the summer monsoon and there is a smaller peak here in the post-monsoon here. So this is the second eigenvector if you look at peninsular clusters away from the west coast. Now, the distribution of the points representing the 37 peninsular stations in the new coordinate space of amplitudes A1-A2.

(Refer Slide Time: 31:23)



You see here so now in the principal components determine only from the peninsular region away from the west coast we get very clearly 7 clusters and these are determined again by the same methodology by looking at seed points and assigning them and so on. So you have 7 clusters that have emerged and what do they look like.

(Refer Slide Time: 31:52)

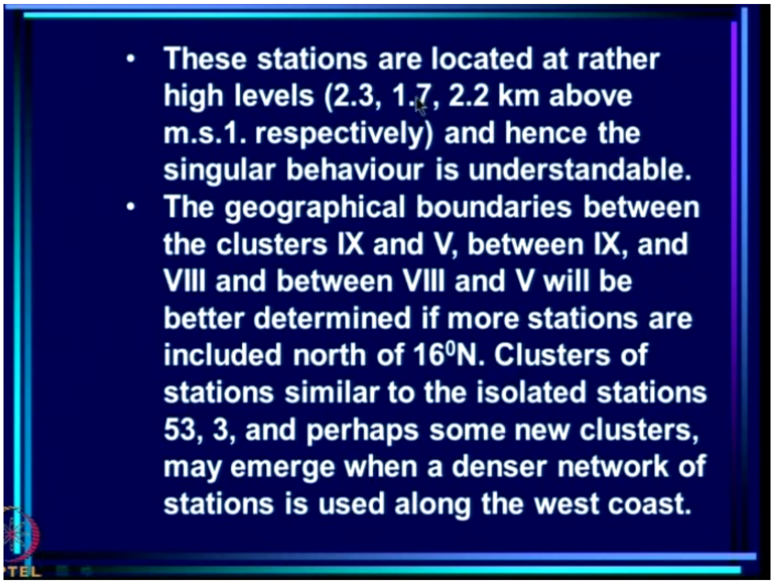


They look like this. So what was originally one big cluster II has now given 7 clusters here and by enlarge there is no ambiguity in assigning stations to any cluster at all except you will see for 3 of them okay. So, 7 clusters 3 to 9 were determined by the method described earlier. The vast majority of stations which is 34 out of 37 can be unambiguously assigned to the different clusters by the method used.

The geographical locations of stations belonging to all clusters identified at the end of the 2-step analysis is shown in the next slide which we have seen and it is seen that the clusters III to IX fall into contiguous areas with the exception of 10, 17, and 18. So we do have some problem in assigning 10, 17, and, 18 which are stations here otherwise you all these stations which you must remember we assign only by looking at their location in the A1, A2 space.

In fact, our occupied regions which are geographically contiguous which is a satisfying thing because we do not expect rainfall patterns to have discontinued it is in space. So by enlarge they are geographically contiguous except for 3 stations.

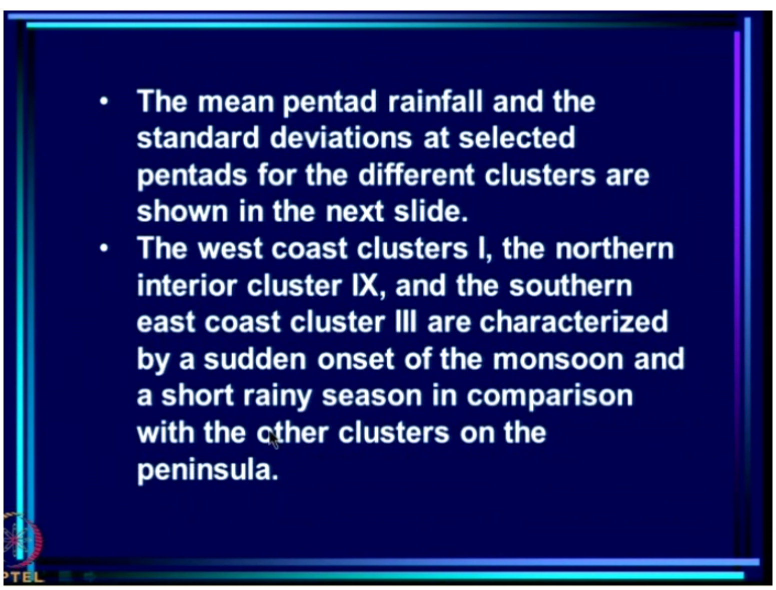
(Refer Slide Time: 33:12)



And these stations in fact happen to be at a higher level 2.3, 1.7, and 2.2 km above the mean sea level. So a singular behaviour is understandable. They will have somewhat different rainfall patterns then the planes around them. Now, the geographical boundaries between the clusters IX and V between IX and VIII and between VIII and V will be better determined if more stations are located, but that is a minor point you can see here. Here there are huge gaps for example.

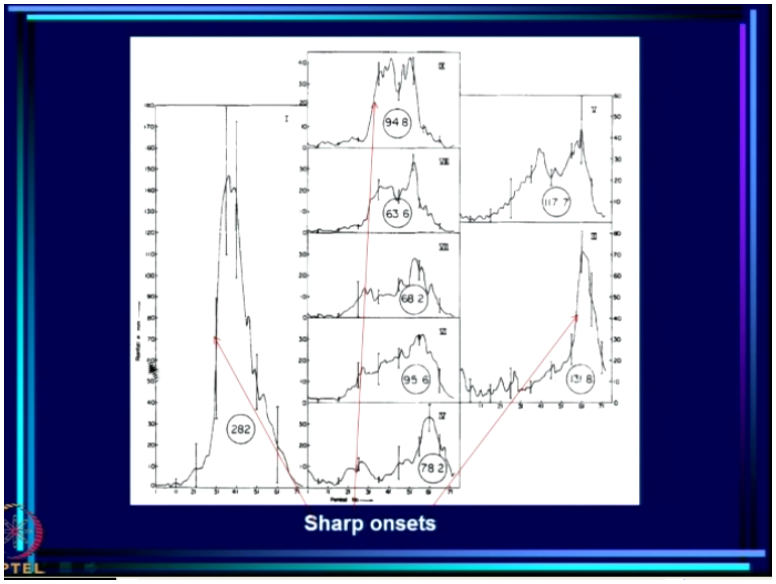
If there were more stations the geographical boundaries would be better determined. Now clusters of the stations similar to the isolated stations 53 and 3 and perhaps some new clusters may emerge.

(Refer Slide Time: 34:00)



If you had a denser network. Now let us look at what the patterns look like so the mean pentad rainfall and the standard deviations at selected pentads for the different clusters are shown in the next slide.

(Refer Slide Time: 34:11)



So this is the west coast cluster which is what came out of the first sorting remember and you see this west coast cluster has a huge very sharp onset and what you see here are mean rainfall it is 282 cm. Then you see very sharp onset here now this is cluster III which is here, this is cluster III. So it involves coastal Tamil Nadu and part of coastal Andhra Pradesh and this also is characterized by a sharp onset that is cluster III here.

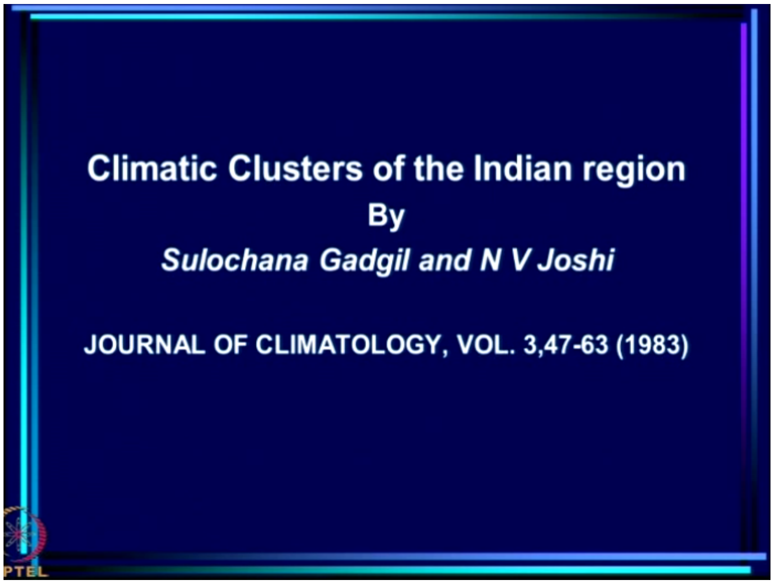
And then we have cluster V which is here which is to the north of this cluster III and that you see both of these have rainfall more than 100 cm for the year as a whole, but the sharpness of the rainy season decreases as you go here to the north. It also decreases as you go the center of the peninsula. These are the central clusters here IV and VI and VII and so on which are here. So IV and VI, and VII, and VIII, and IX.

So these are the central clusters here going along the central longitude and you find that southern most point has a very clear peak in the October, November season like Chennai does or like the east coast, south east coast of peninsula does, but what happens as you go northward is that the rain pattern becomes somewhat broader and as you go to the north in fact you have seen that the peak in October, November has totally disappeared and it is dominated by the summer monsoon.

So you see that there is considerable variation in the patterns across the peninsula and as I mentioned there are sharp onsets and I will come back to this point both for cluster I which is the west coast for cluster III which is the east coast of southern peninsula and also for cluster IX which we can see here. So this cluster also has a very sharp onset. Relative to that these others are somewhat gentle onsets only these 3 stations.

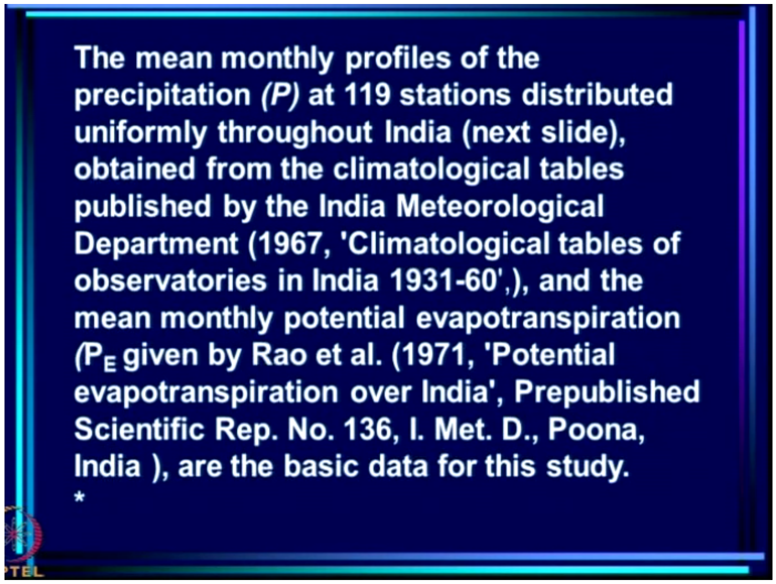
These 3 clusters have sharp onsets and as I said I will return to the point of determining the onset dates towards the end of the talk.

(Refer Slide Time: 36:51)



Now so far we have been talking of peninsular clusters, but we would like to have clusters for the entire region which we have also determined.

(Refer Slide Time: 37:06)



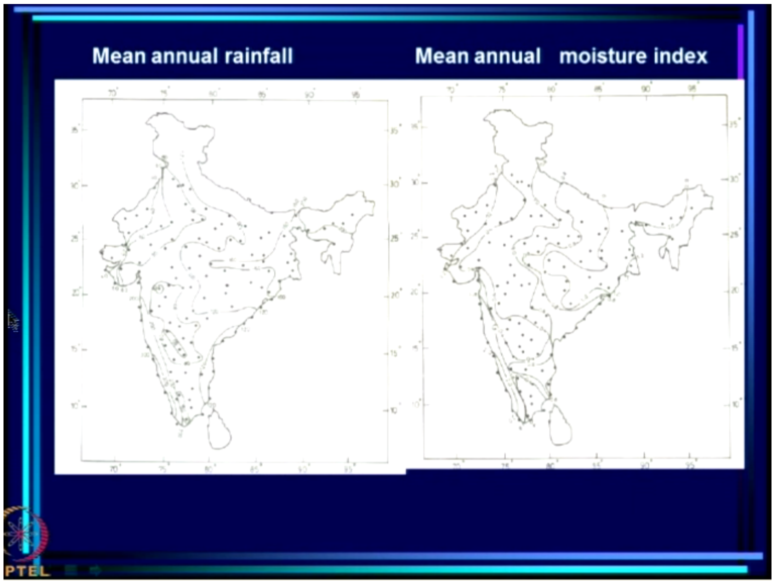
And this we did by taking 119 stations distributed uniformly throughout India.

(Refer Slide Time: 37:10)

You can see reasonably good distribution throughout India and what we did in this case was what only did we analyze the rainfall which is P precipitation, but we also got the data on mean monthly potential evaporation, evapotranspiration. See this is required by agricultural scientist because P over PE gives the moisture index. How much is the precipitation relative to the potential evapotranspiration?

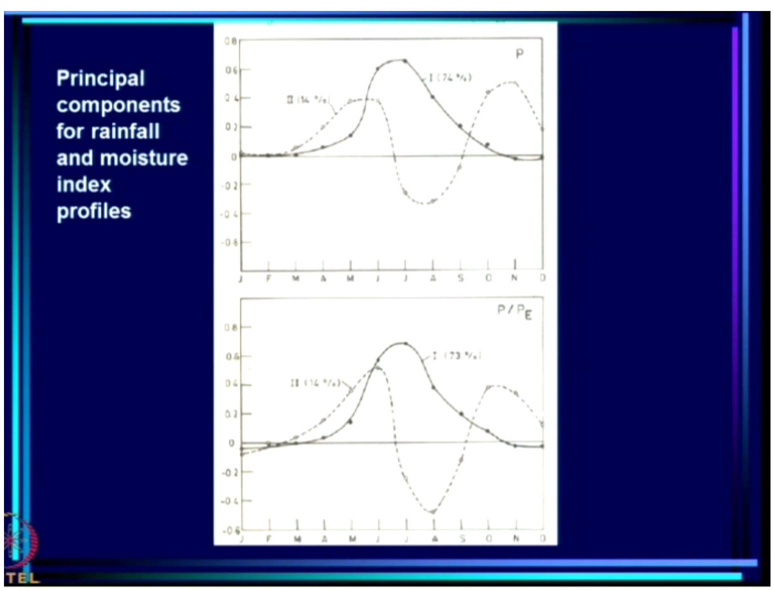
So we looked at the moisture index also and in that paper we have also looked at temperature profiles, but I am not going to discuss them here okay. So, this is the mean annual rainfall over.

(Refer Slide Time: 37:56)



The region considered and we have seen patterns like this before. You have high rainfall over the west coast which decreases here and little more over the east coast here and then this is the monsoon zone in which the rainfall decreases as you go this way. The moisture index annual patterns look very similar to rainfall very similar because the potential evapotranspiration does not vary too much.

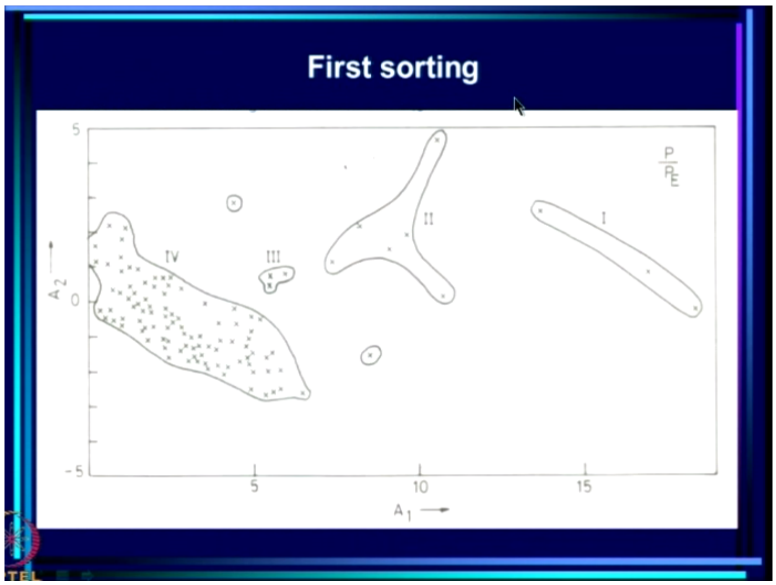
(Refer Slide Time: 38:28)



So, it is rainfall that determines the moisture index. Now remember we have gone to the monthly scale for this study. So this is again the first principal component. We use similar method now to reduce the dimensionality from 12 to 2. The first principal component has explained 74% of the variance. The second one has explained also a substantive amount 14% so together they explain 88% of the variance.

And this is for rainfall and this is for moisture index again you see that the first component is very similar summer monsoon and the second is pre- and post-monsoon. Very similar to what we had seen for the peninsular region.

(Refer Slide Time: 39:13)



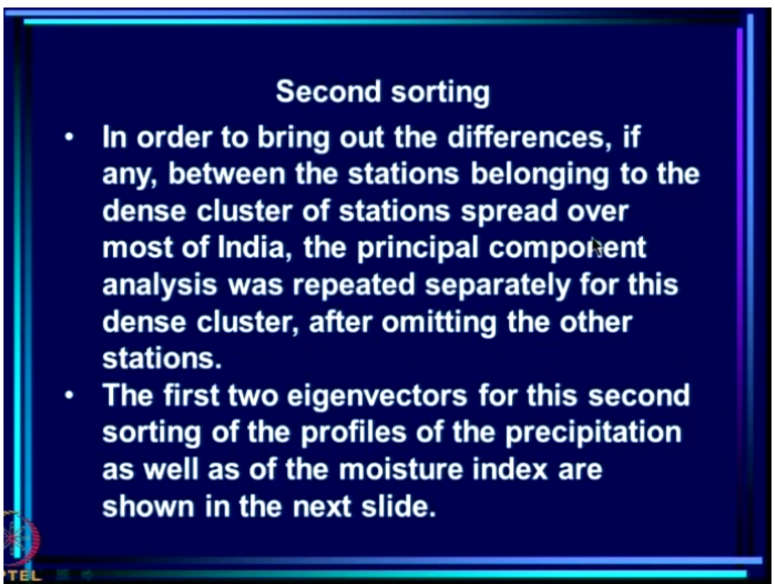
Now having done this we then repeat the exercise having determined the components we determine the amplitudes for each station and what we get from the first sorting then is this is A1 and A2 and these are all the stations over the entire Indian region so there is 1 cluster here, I then another cluster II, then there are some isolated points and a small cluster III here and a big densely packed cluster IV and how do they look on a map.

Well this most of India is the densely packed cluster IV. Now I and II are simply regions that are shown here see I is the highest rain part of the west coast which is here and II is the entire west coast other than the highest one. Similarly, this part of north east is also II and in here is cluster III. So I, II, and III are all heavy rainfall regions and the patterns sorted out are for this heavy rainfall regions and the rest of India is left as I cluster 1 single cluster.

And these are the mean patterns of the clusters then you can see that cluster I which has the heaviest rain is here again very similar this we have seen is it is very similar to the cluster I of the peninsula. This is cluster II which is both north east and part of west coast then we have cluster III which is a little bit of part of north east.

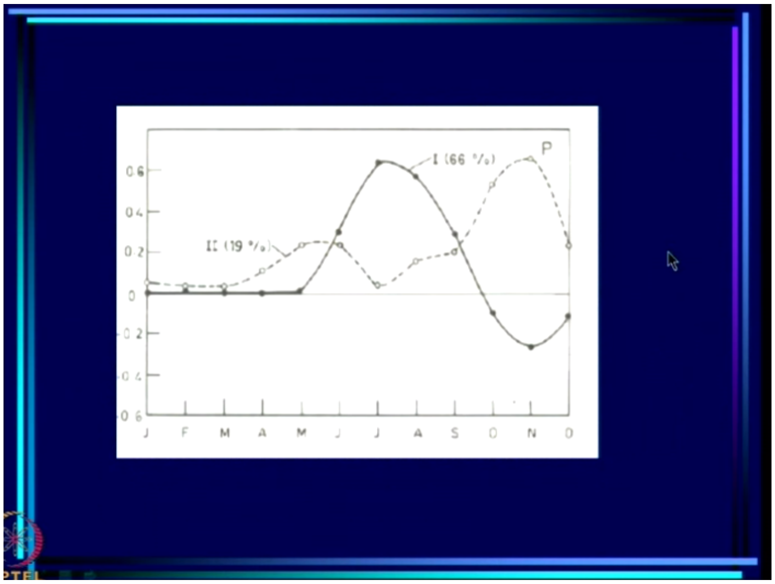
And then we have cluster IV which is the entire rest of India and you can see that in cluster IV unlike the other ones you know the variation from station to station is huge compared to the mean and this is suggesting that one should do one more sorting of this. See the variation between these stations was overwhelmed by the variation between the high rainy region and the not so high rainy region and that is why appeared as one cluster in the first sorting.

(Refer Slide Time: 41:30)



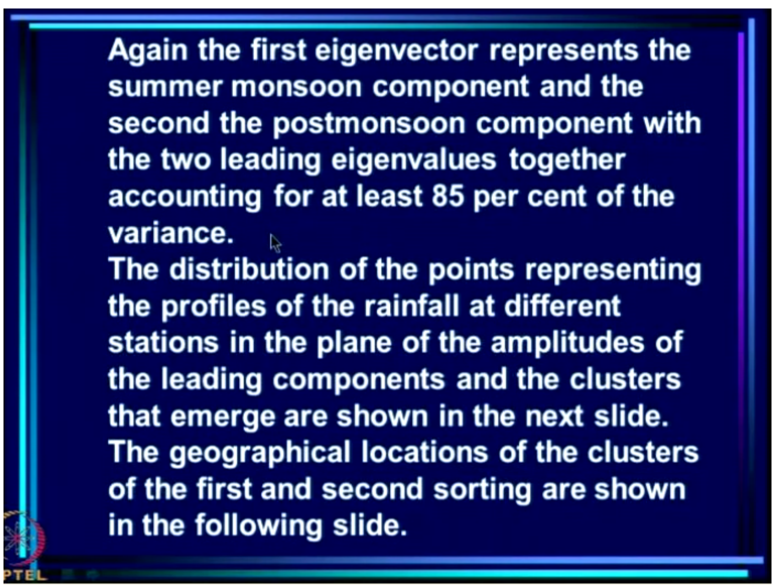
But, we again bring out the differences between the stations belonging to the dense cluster of stations spread over most of India by doing a second sorting do again a principal component analysis only for these stations after omitting the other station and the first 2 eigenvectors for this sorting of the profiles are given here.

(Refer Slide Time: 41:51)



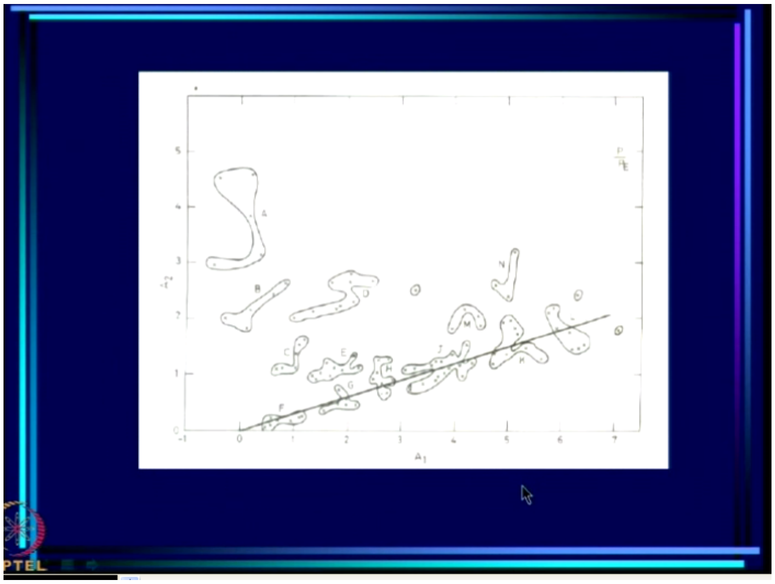
Now again the first eigenvector turns out to be the summer monsoon component in this case and with the big negative values here and the second component is the post-monsoon and pre-monsoon component.

(Refer Slide Time: 42:06)



So, again the first eigenvector represents the summer monsoon component and the second post-monsoon component with the 2 leading eigenvalues together accounting for at least 85% of the variance. The distribution of the points representing the rainfall profiles at different stations in the plane of the amplitude are shown here.

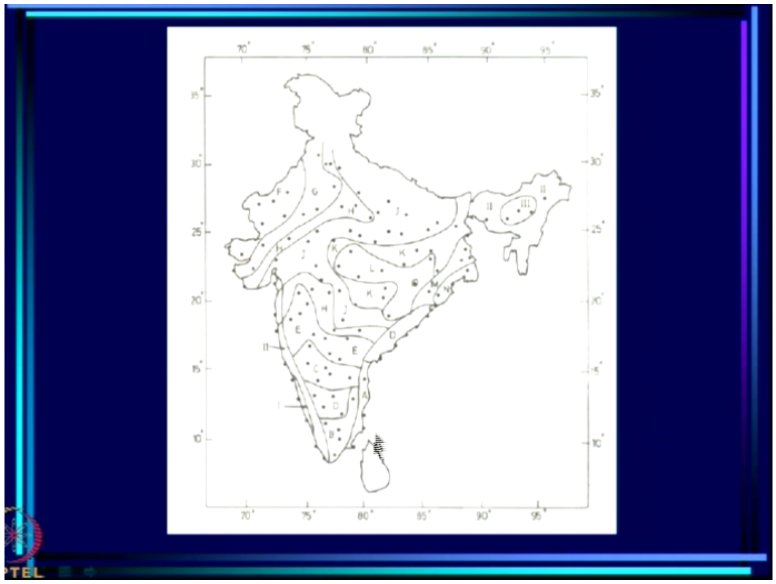
(Refer Slide Time: 42:25)



And again by the method I am not going to repeat that by the method that I described earlier we determine the clusters here and notice these are genuine clusters with clear gaps between the clusters and what you see these are clusters A, B, C, D, which we will see the locations on the map, but what is interesting is these set of clusters from F onwards all happened to be on the same straight line here.

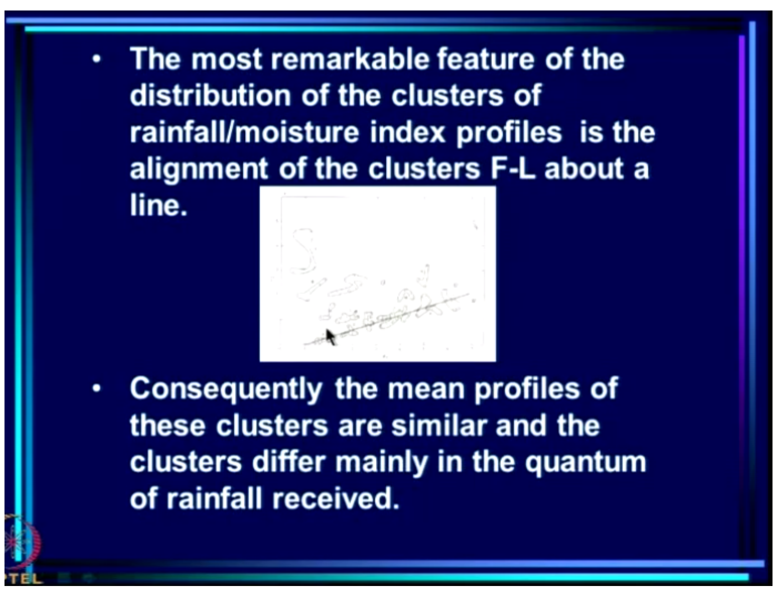
See a straight line passes through all of them that is saying that A2 and A1 are proportional to one another along the straight line and so, these are the clusters that we see here. You saw first these are A and D. These are the clusters that you had seen here that we just point out which these are. This is A, B, and D and this is C. So B and D have roughly same A2, but increasing value of the summer monsoon component and A has relatively less value of the summer monsoon component.

(Refer Slide Time: 43:43)



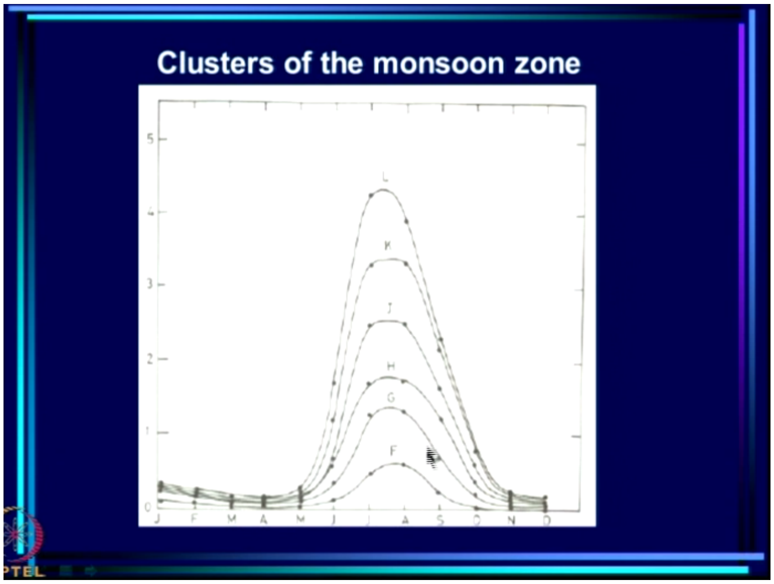
So you have A is here. These are the minimum value of the summer monsoon component then we have D here, then B, D, C, E as you go up so this is B, D, C, and E. So as you go up in the peninsula and then we come to the most interesting part which is F, G, H, J, K, L, M, N. So these are all along the monsoon zone. Remember the monsoon zone is around this region here. So these are all along the monsoon zone going from the minimum rainfall here to the maximum rainfall here.

(Refer Slide Time: 44:37)



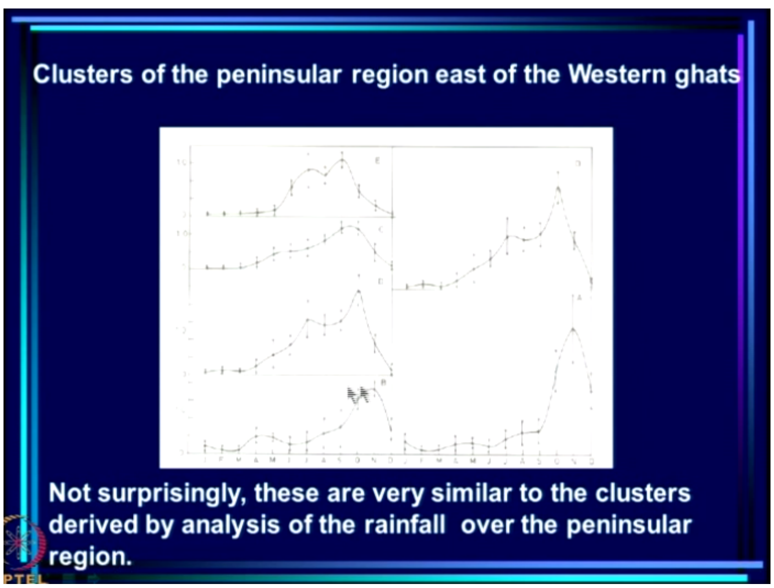
And what you see is that see these are all along the this is the most remarkable feature from F to L all of these are along the single line the most remarkable feature of the distribution of clusters of rainfall and moisture index when we look at the entire Indian region is the alignment of the cluster about a line. Consequently, the mean profiles of these clusters are similar and the clusters differ mainly in the quantum of rainfall receives.

(Refer Slide Time: 45:00)



So F is the one that is in the northwestern part and as you go along the monsoon zone to the east you will find the rainfall increases, but the pattern is pretty much the same.

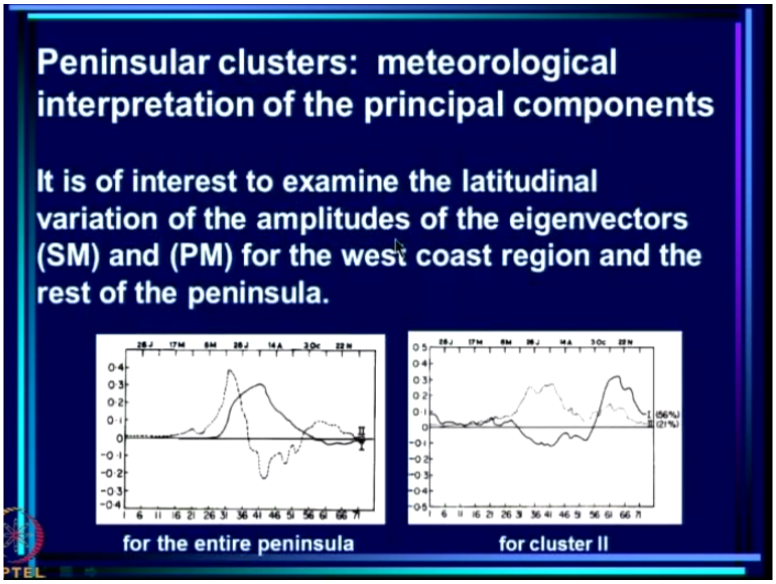
(Refer Slide Time: 45:15)



On the other hand, if you look at the peninsular clusters and remember this E I am sorry A is the south eastern part. This is corresponding to the same cluster that we had found in the peninsular region and this is the northern part of the east coast and this again is going along the central longitude.

So not surprisingly the peninsular clusters come out very much same as we had determined before what is interesting is what happens to the clusters in the northern part and there we are getting in fact mean pattern is the same so it is the same single system that is giving it rainfall.

(Refer Slide Time: 45:53)



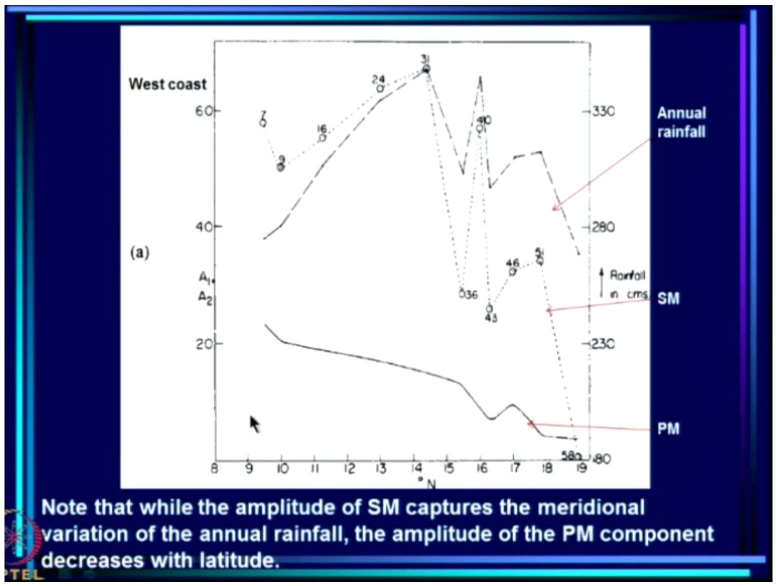
Now so far so good we have now come up with the final homogenous regions are climatic clusters if you wish of the rainfall patterns and so if within this region you would say that the rainfall profiles are similar; similarly, within any of these regions you can say that the rainfall patterns are similar.

These are natural groups which are homogenous relative to rainfall patterns which is what we had sorted out to actually find in the first place. So this is fine. These are the actually come up with the climatic clusters of the Indian region using mean rainfall patterns and the once using mean moisture index are almost identical to this. Now and this remember is the pattern of variation in the monsoon zone.

Now can we having determined these principal components can we interpret what do these principal components mean they are weightages to different pentads so they represent some kind of a profile, temporal profile and what does it mean so it is of interest to examine the latitudinal variation of the amplitudes of the eigenvectors SM and PM for the west coast region and the rest of the peninsula.

So, we have here in fact rest of the peninsula we have this is the main component the PM component this summer monsoon is the second component, whereas over the west coast this is the first component and the second component is pre- and post-monsoon.

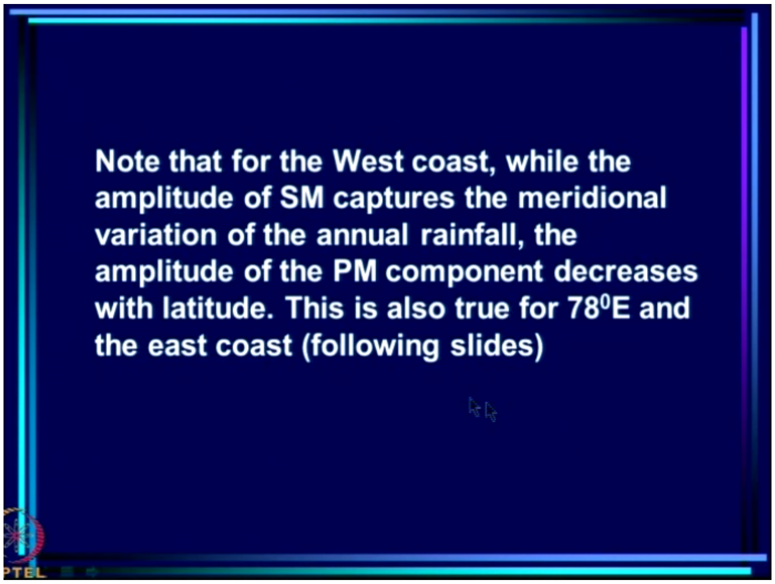
(Refer Slide Time: 47:40)



So let us look at how the amplitudes vary with latitudes. So, first look at the west coast. What you see here is the annual rainfall and this is the summer monsoon component amplitude and this is the post-monsoon amplitude. What is interesting is and here it is not surprised that in fact the annual rainfall variations with latitude are very similar to the variations of the amplitude of the summer monsoon component. So this is very, very reasonable.

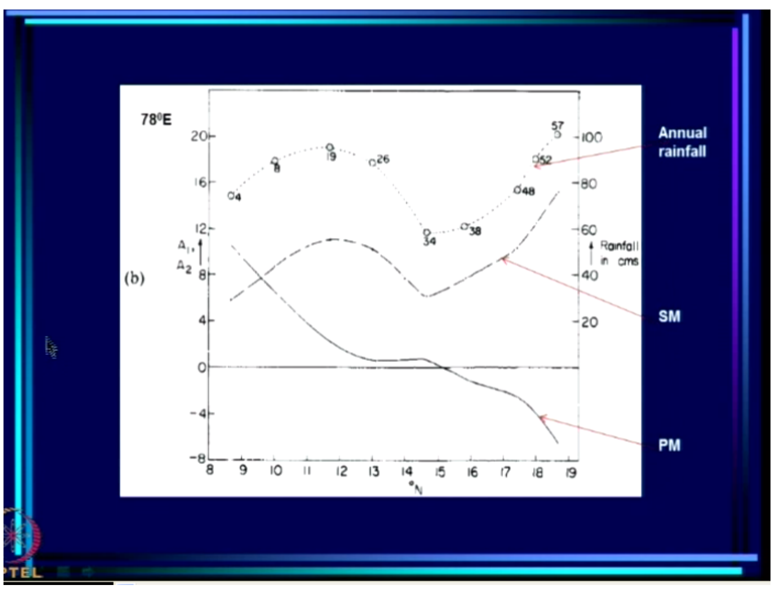
So how does the pre- and post-monsoon component vary? It actually the amplitude decreases as you go north. See we are at 9 degrees here and we are going north up to about 19 or 20 this is latitude of close to Bombay. So what is happening is that the second mode amplitude is decreasing continuously.

(Refer Slide Time: 48:35)



Now let us note that for the west coast while the amplitude SM captures the meridional variation of the annual rainfall the amplitude of the PM component decreases with latitude. What is interesting is?

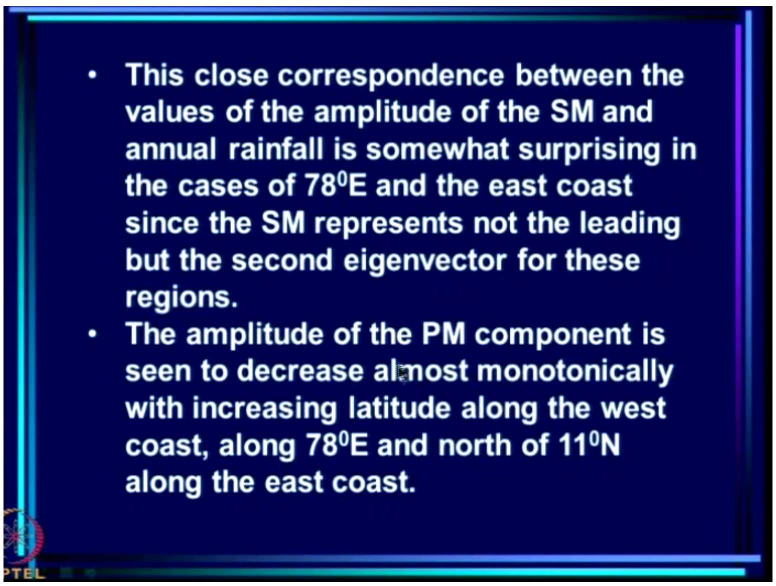
(Refer Slide Time: 48:46)



This PM amplitude decreasing with latitude is also true for 78 degrees east remember 78 degrees east is the central longitude here so this is where this is 75 and this is 80. So 78 degrees east actually cuts across this 78 east is going across like this and we are going only up to peninsula. So it is the central longitude of peninsula if you wish. Now if you go to 78 east again interestingly, the annual rainfall variation is captured by amplitude variation of the summer monsoon component.

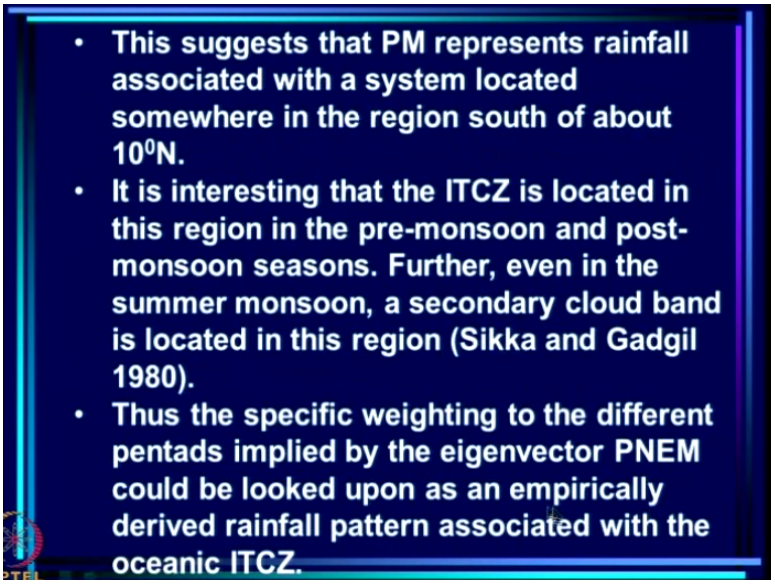
This is very interesting because actually here the dominant component is not the summer monsoon one, but the post-monsoon month pre and post-monsoon and what you see again is a decrease in amplitude of the PM component as you go towards the north. This is very interesting so and now what happens on the east coast. East coast also similar thing, east coast also to some extent the variation the nature of variation of annual rainfall is determined by the amplitude of SM and PM on the other hand simply decreases with latitude which is very interesting.

(Refer Slide Time: 50:11)



So close correspondence between the values of the amplitude of the SM and the annual rainfall is somewhat surprising in the case of 78 degrees east and the east coast since SM represents not the leading, but the second eigenvector so we have to understand how this is has come about. The amplitude of the PM component is seen to decrease almost monotonically with increasing latitude along the west coast, along the central longitude of the peninsula that is 78 and north of 11 along the east coast.

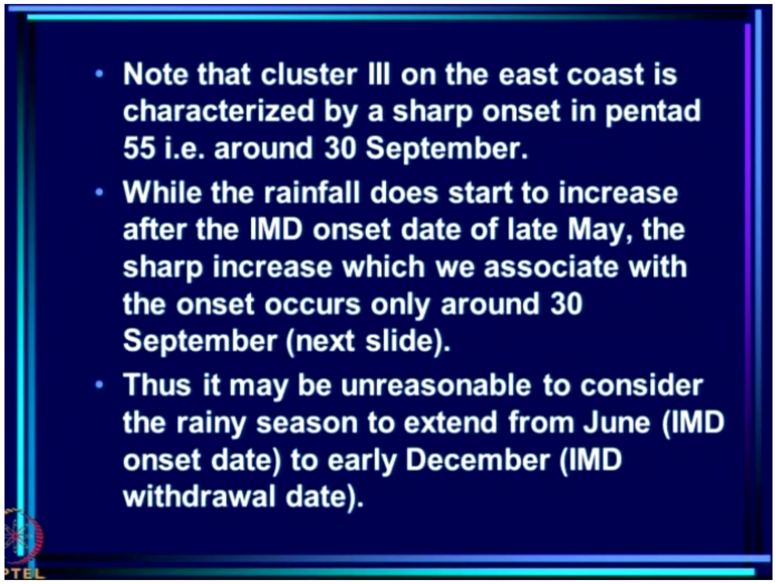
(Refer Slide Time: 50:40)



This suggests that PM represents rainfall associated with a system located somewhere in the region south of about 10 degrees north. It is interesting that the ITCZ is located in the region in the pre-monsoon and post-monsoon season. Further, even in the summer monsoon a secondary cloud band is located in this region. Thus the specific weighting to the different pentads implied by the eigenvector PM could be looked upon as an empirically derived rainfall pattern associated with the oceanic ITCZ.

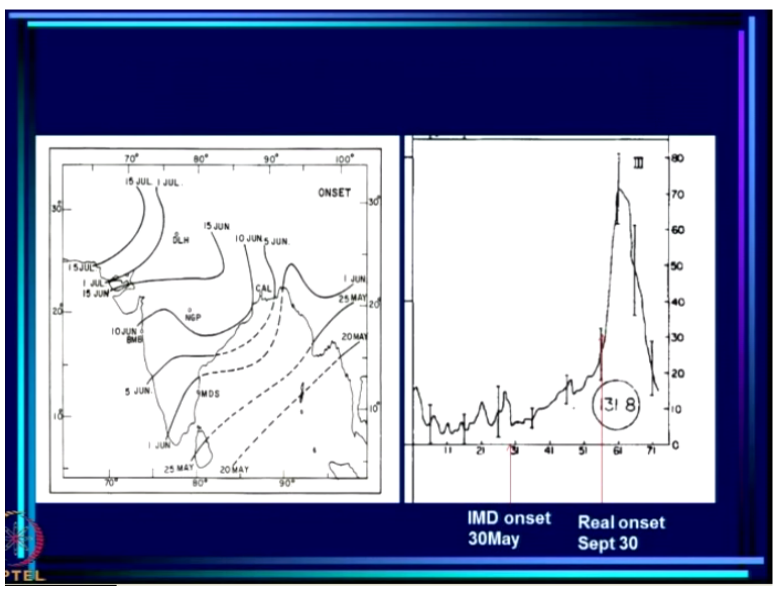
So this is a very interesting interpretation that we have that if we did not have the continental ITCZ at all had only the oceanic ITCZ you would expect the rainfall pattern like the PM that we have determined. Now what about the implication for onset dates. We began this lecture by looking at the onset dates given by IMD.

(Refer Slide Time: 51:35)



And in particular, you may recall that this zone III which involves which is the southern part of the east coast of peninsula we notice that cluster III is characterized by sharp onset in pentad 55 around 30th September while the rainfall does start.

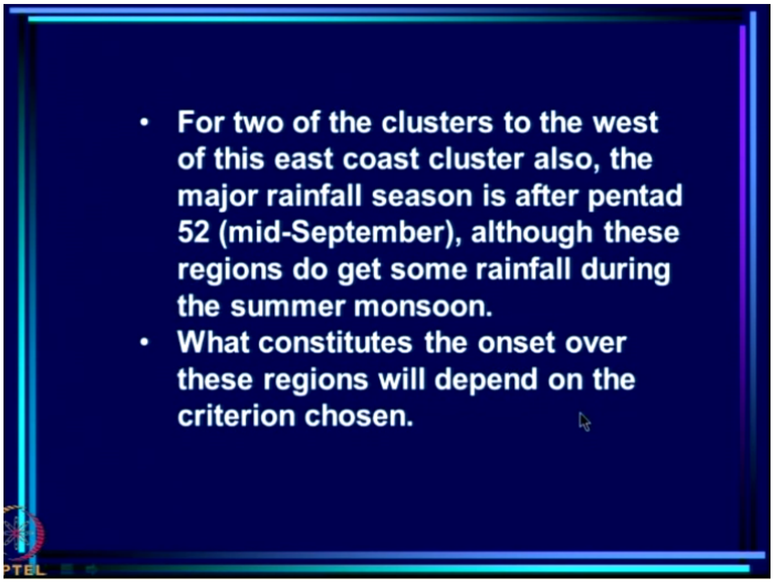
(Refer Slide Time: 52:06)



So we look at this cluster III which is on the east coast and what we find is that in fact there is a sharp onset what you call onset of the rainy season occurs only in September 30, but if you look at the IMD chart the IMD onset occurs on 30th of May itself. After that there is an increase in rainfall but very gentle, and the sharp increase only occurs in on 30th of September. So while the rainfall does start to increase after the IMD onset date of late May the sharp increase which we associate with the onset occurs only around 30th September.

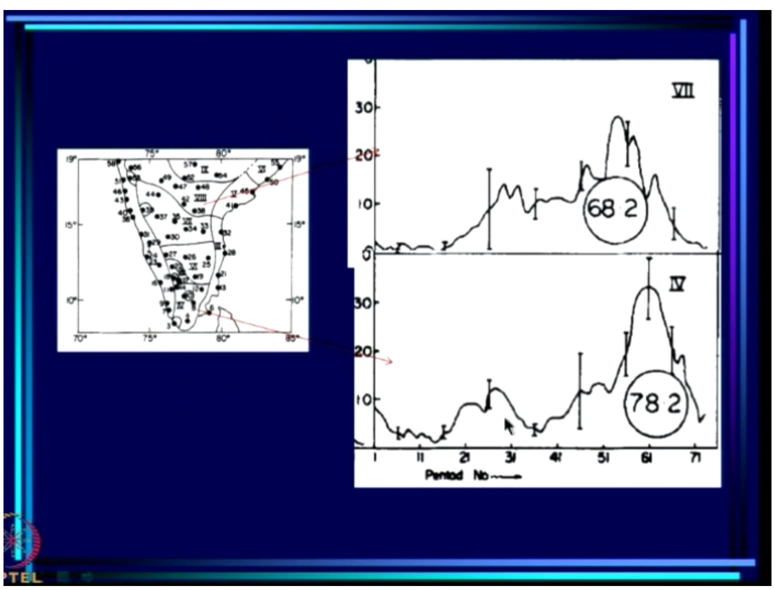
So it may be unreasonable to consider the rainy season to extend from June to early December and it may be better to consider it from end of September or so.

(Refer Slide Time: 52:49)



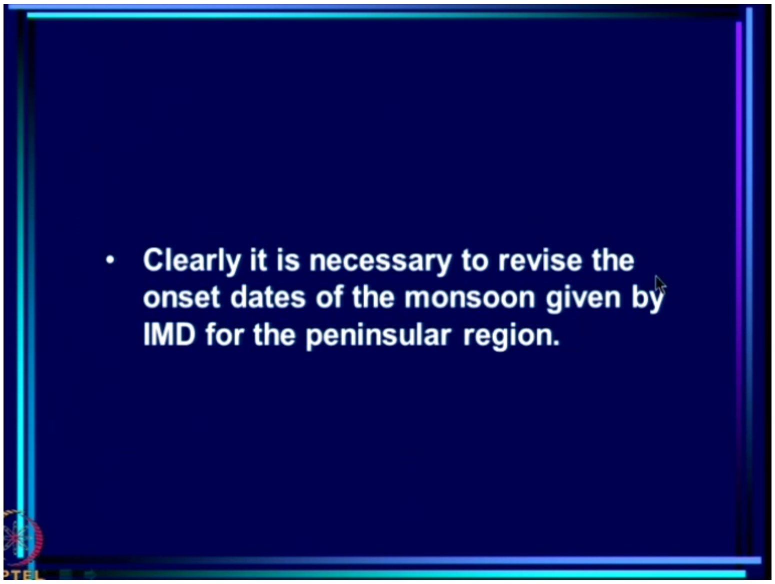
Now similarly, for the 2 clusters to the west of the east coast also the major rainfall season is after pentad 52 that is after mid-September.

(Refer Slide Time: 52:56)



See the major rainfall season is this one here. Certainly for cluster IV it is here and even for this one it is after this is the major rainy seasons after pentad 52 and although these regions do get some rainfall during the summer monsoon. What constitutes the onset over these regions will depend on the criterion chosen because it is not a very sharp onset, but you can see very clearly that the major rainy season is around here.

(Refer Slide Time: 53:31)



So, clearly it is necessary to revise the onset dates of the monsoon given by IMD for the peninsular region.