Machine Learning for Earth System Sciences
Prof. Adway Mitra
Department of Computer Science and Engineering
Centre of Excellence in Artificial Intelligence
Indian Institute of Technology, Kharagpur

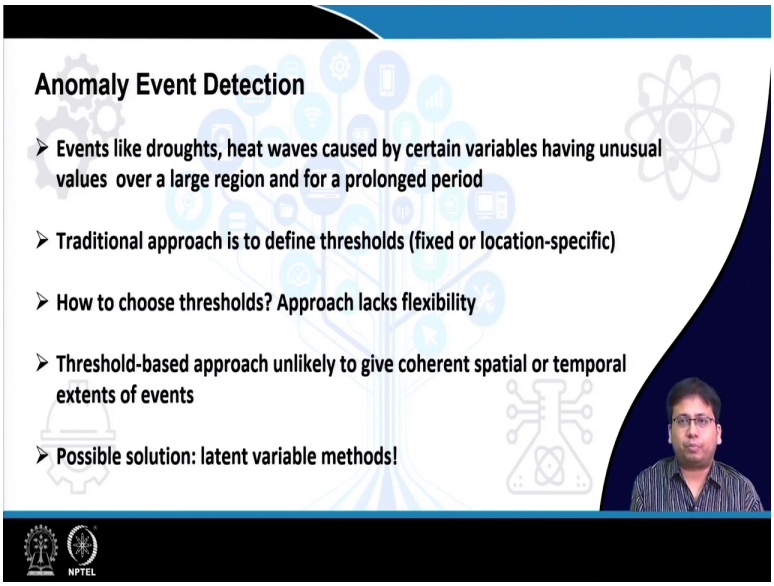
Module - 03
Machine Learning for Discovering New Insights
Lecture - 18
Detection of Anomaly and Extreme Events

Hello everyone, welcome to lecture 18 of this course on Machine Learning for Earth System Sciences, we are in module 3 which deals with the applications of Machine Learning in Discovering New Insights in the Earth Systems. So, the topic of today's lecture is the Detection of Anomaly and Extreme Events using Machine Learning.

(Refer Slide Time: 00:47)

So, the concepts we are going to cover today are; anomaly events modeled using Markov random field, remember we have discussed about Markov random fields in one of the lectures in module 2 and we will also discuss about identification of extreme events using computer vision concepts.

(Refer Slide Time: 01:06)



So, first of all let us talk about anomaly event. So, events like droughts, heat waves etcetera which are caused by certain variables having unusual values over a large region and for a prolonged period of time. So, traditional approach to identify such events is by defining thresholds which might be fixed or universal or location specific and then identifying like when and where though the variable in question exceeded those thresholds.

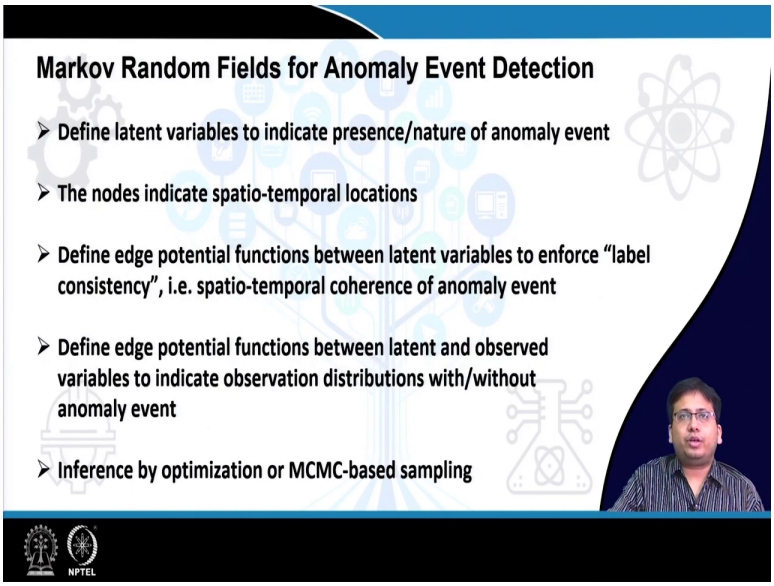
But, the question then arises about how to choose these thresholds? So, that approach always lacks flexibility. I mean whichever threshold you choose; however, you choose it will never be like it will always be debatable and besides the threshold based approach is unlikely to give coherent spatial or temporal extents of the events.

That is suppose you want to identify that these are the precise or this is the precise region where the heat wave is present at any given time or this is the period over which the heat wave was active or this is how the heat wave travelled from one region to another. It is very difficult to get it using a threshold-based approach, because like we will see that one region here is above the threshold.

So, like that is having a heat wave, but a neighboring region there it might be slightly below the threshold so, it is not having the heat wave and so on. So, whatever threshold you choose unless you choose a very low threshold you are unable to like identify such coherent regions.

But then again if you choose a low threshold that defeats the whole purpose of, because we are usually interested in significant deviations from the normal or in the long period average. So, these kind of threshold based approach like this point we have already discussed in another lecture in module 1 also. So, the possible approach is the that of the latent variables.

(Refer Slide Time: 03:09)



So, we might define latent variables like to indicate the presence or the presence or the nature of the anomaly event, that is whether an anomaly event is there or secondly, what kind of it, like is it a positive anomaly or negative anomaly, then again among positive anomalies, there can be different types of anomalies, I mean like is it a drought or is it a heat wave or is it a like what or is it a cold wave so, like it different kinds of like things are there.

So, like we can use these latent variables which are probably discrete to like specify the either the presence or the nature of the anomaly event. And these variables these will have to be spatio-temporal. So, we represent the whole thing as a graphical model as discussed in one of the earlier lectures. So, like each nodes for they will correspond to a random variable which might be

either the observations which we treat as the observed random variables or these latent variables and the each nodes also indicate some kind of a spatio-temporal location.

So, now we can define edges between neighboring locations or I mean that is I mean neighboring nodes, neighboring either spatially or temporally and we can define the edge potential functions between the latent variables to enforce the "label consistency" that is to say the spatio-temporal coherence of anomaly events, which basically means that if a particular location is having some kind of an anomaly event its neighboring location are also likely to have the same event or if today in this location there is an anomaly event it is likely that tomorrow also it will be there.

So, that is the spatial and temporal consistency coherence which can also be considered as "label consistency", because these latent variables as I mentioned that these are discrete. So, they can these latent variables can be called as a labels. And so, we besides we also can define the edge potential functions between the latent and observed variables. So, like there are two kinds of edges one is between the latent variables and the other is between the a latent variable and an observed variable.

And so, these the second kind of edges their edge potential functions will indicate the observation distribution in the presence or absence of the anomaly event. So, like if the anomaly event is there then of course, the variable of interest it will have some kind of values and in the absence of the anomaly it will have a different kind of values. So, that is kind of self-explanatory, I mean we know that say suppose we are considering heat waves in that case we are we will be focusing on the temperature variable.

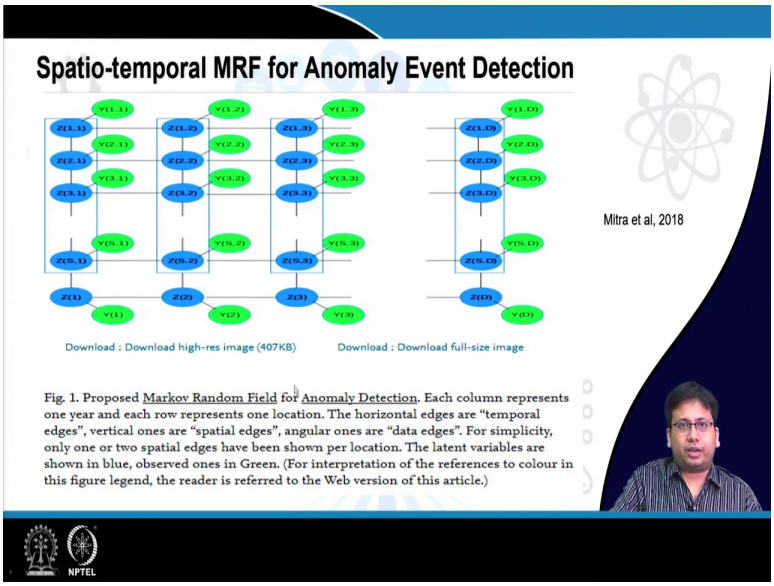
So, we can understand that, if the heat wave is present then the temperature will have like generally high values. So, it will be represented by some kind of a probability distribution whose mean may be high, but if the heat wave is not there it can take a different distribution with lower values. So, in that case again the mean of the expectation or mode of that distribution might be a lower value.

So, in the presence or absence of anomalies or in the presence of different types of anomalies the variables can have the observed variables they can have different distributions. So, that

distribution can be specified using these edge potential functions between the latent variable and the observed variable.

So, the latent variables as well as the different parameters of these potential functions, these are the things we do not observe, these are the things which we have to infer or estimate while the things we observe are these values of the observations or the observed variables. So, this inference like we can do either by optimization or by MCMC based sampling. So, remember in lecture 15 we have already discussed how this thing that is we have already discussed Gibbs Sampling, variational inference, etcetera how that can be done.

(Refer Slide Time: 07:27)



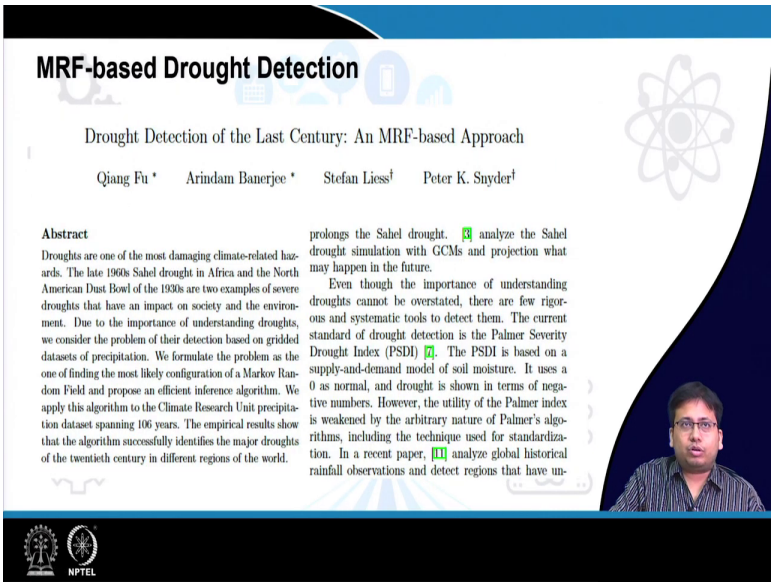
So, like this is, what our spatio-temporal graphical model may looks like. So, let us say that these blue variables these are the hidden or the state variables and the green nodes these are the observations. So, now you can see that these columns or these vertical boxes these correspond to time points. So, and these rows they correspond to spatial locations.

So, like you can see that this they the they are named as or indexed as $Z(1, 2)$ etcetera. So, the like one of the indices is for space the other one is for time and so, these vertical arrows these are the spatial edges they connect two nodes which belong to neighboring locations. And these

horizontal edges these are the temporal edges which connect at the nodes at the same location, but at different points of times.

And then there are these diagonal edges these are the data edges which connect the spatio-temporal variable at any given time location and time with the corresponding observation. And so, like each of those edges have the edge potential function as in the same manner which I defined. So, these green nodes these are the these are observed we know their values, but the blue node we do not know their values.

(Refer Slide Time: 08:56)



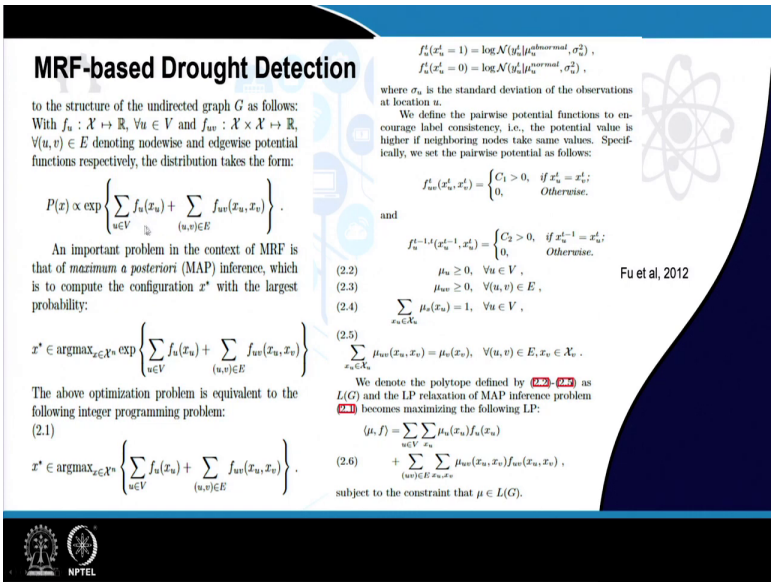
Now, you like using this framework I will briefly talk about two case studies; one is for Markov Random Field based Drought Detection in the 21st century. So, I mean the 20th century. So, the droughts are one of the most damaging climate related hazards. The late 1960s Sahel drought in Africa and the North American Dust Bowl of the 1930s are two examples of severe droughts that have an impact on society and environment. Due to the importance of understanding droughts we consider the problem of their detection based on gridded datasets of precipitation.

So, in this case like in the case of drought the variable of interest is precipitation instead of that if we had heat waves then it would have been say temperature, may be daily maximum temperature or daily minimum temperature along with maybe humidity etcetera. We formulate the problem as

the one of finding the most likely configuration of a Markov random field and propose a efficient inference algorithm. We apply this algorithm to the Climate Research Unit precipitation data set that spanned 106 years.

The empirical results show that the algorithm successfully identifies the major droughts of the twentieth century in different regions of the world.

(Refer Slide Time: 10:16)



So, this is like this is the Markov random field. So, this is what the joint so, as I said the Markov random field what it basically does is like it specifies a joint distribution over all the variables. So, there are two kinds of variables: the hidden variables as well as the observed variables. So, the whole thing the joint distribution is like the basically the product of the all the click potentials.

So, here we can in this kind of a graph we have two kinds of clicks one is the we like we can have the individual nodes that is clicks of size 1 or the edges which are the clicks of size 2. Of course we need not have them separately the because the each node is anyway part of an edge except for the isolated node. So, like we can say that except for the isolated nodes like we the remaining nodes which are part of the an edge their potentials can be absorbed into the edge potential.

So, we will but whatever it is. So, what these $x_u's$, $x_v's$ etcetera, these are the values of the those variables, but while the values of the observed variables are known those of the hidden variables are unknown.

So, the task is we need to choose the values of the hidden variables in such a way that this probability the joint probability is maximized. So, this is posed as some kind of an optimization problem where the task is to calculate the x in such a way, this x is like the collection of all the variable values including the hidden and the observed. So, we have to choose x in such a way will that it this thing is this joint this probability is maximized.

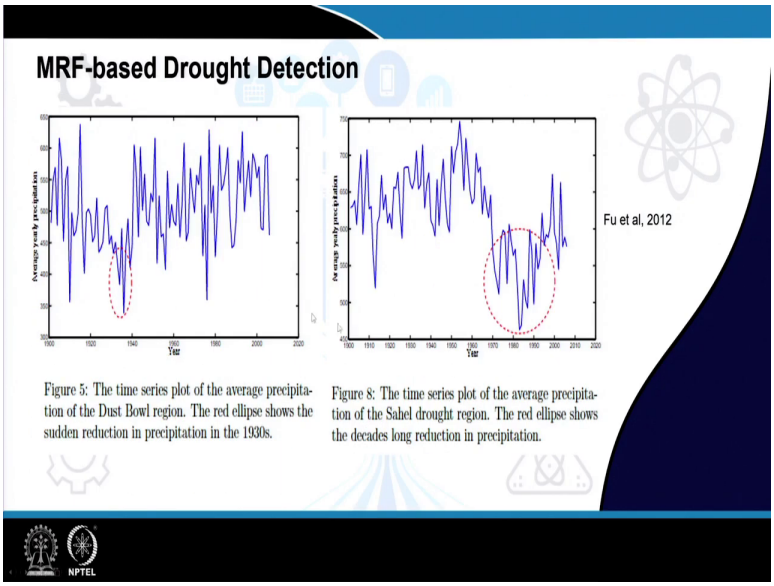
So, they in this paper they formulate that as a something like a linear as some kind of a this LP problem as a linear programming problem and they solve it. So, they define the node and edge potential in this way. So, these are like you can say these are the data edge potential. So, here as I said the like these the potential functions on these diagonal edges connecting the latent state to the observation.

So, like so, in both cases the distribution is considered to be a Gaussian distribution. So, if it is like if it is normal then the mean is μ we say it is μ^{normal} which may vary from one location to another and if it is if an anomaly event is actually happening then it is the distribution is called $\mu^{abnormal}$ its parameters, again this parameter may also vary from one location to another.

And then there are these potential functions for this spatial and temporal edges so, they are chosen in such a way that this spatio-temporal coherence is maintained that is whenever the variable the anomaly variable at two neighboring locations are the same either both are having the anomaly or both are not having the anomaly the corresponding edge potential function takes a high value, but if they are different then it takes a low value.

So, the reason for defining this potential function is that we want to promote this level consistency of spatio-temporal coherence. So, that I mean when whenever we have these kinds of things the overall probability of that particular configuration is increased. So, similarly we can have that on the temporal edges also we have the same at the temporal coherence constant. So, use it by defining the edge potential functions in this way we like we formulate this problem and solve it using linear programming.

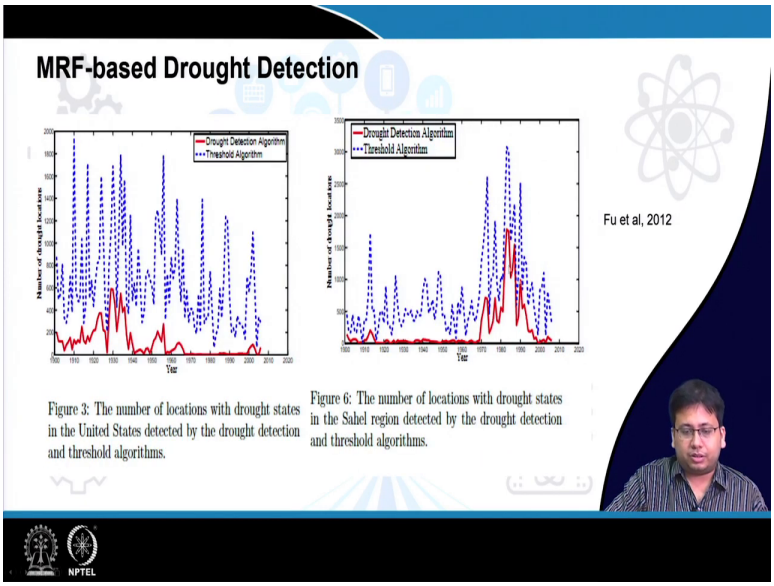
(Refer Slide Time: 14:15)



So, these are some of the results. So, like as you can see if you have the time series of precipitation in the Dust Bowl region of America in this case and the Sahel region of Africa in this case. So, here you can see that there is a particular period of time when the precipitation value is quite low, here also there is such a period, but the it is such that like if you put some kind of a threshold you will not be able to identify that this region so well, because like even within this period there are one or two exceptional years in which that threshold is exceeded.

So, if you put some and the only way to identify is to put a high threshold like this, but then the like a lot of these isolated years they will also be considered to be droughts which I do not want. I am only interested in such droughts which extended for a considerable period of time even though within that period not every single year might have been below any particular threshold.

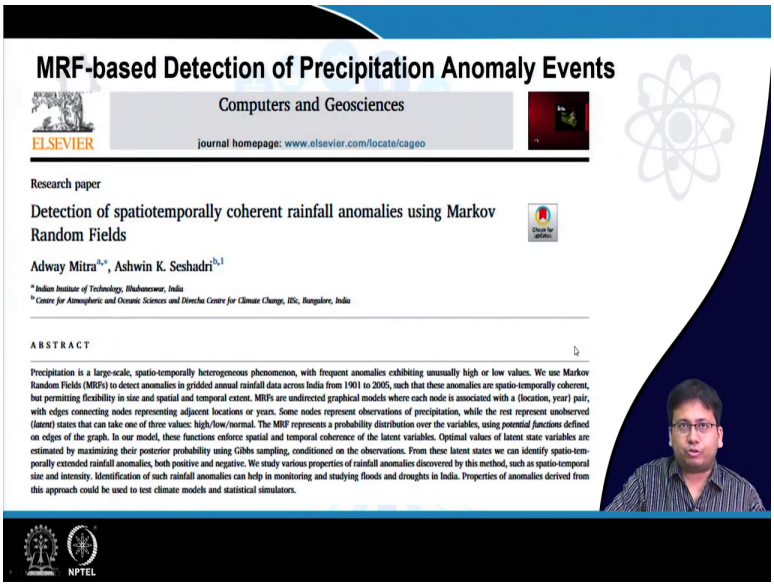
(Refer Slide Time: 15:30)



So, to like to get that kind of flexibility we prefer the Markov Random Field approach over the threshold approach and this is the what happens when we go for the threshold approach we get a large number of these spurious droughts. So, here they have plotted the number of drought locations in the using the threshold as well as in the proposed MRF based algorithm.

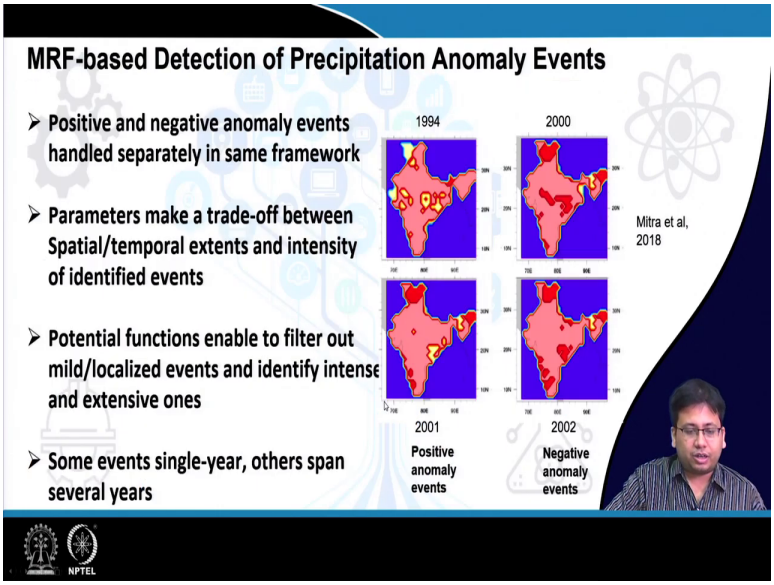
So, here as you can see the threshold approach it gives a lots of lots of spurious locations as droughts while in case of MRF it is like the number of locations is much less and it can be verified by experts I mean the experts will have their own definition of droughts which is based on a number of parameters. So, they can actually corroborate that it is this red series which is much closer to what is like their truth.

(Refer Slide Time: 16:18)



The similar approach was also used to detect coherent spatio - temporal rainfall anomalies over the Indian region. So, in this case like there were like, unlike here we are considering only one type of anomaly that is droughts, here we are considering two kinds that is both positive as well as negative anomaly.

(Refer Slide Time: 16:46)

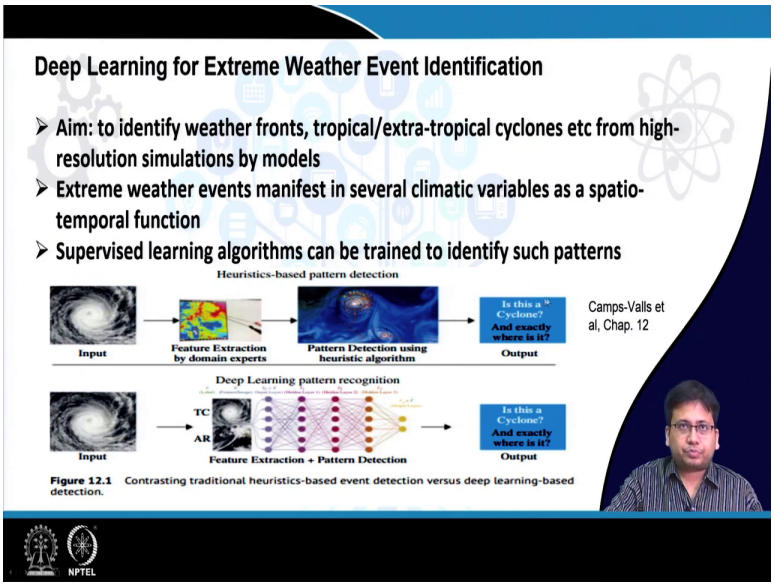


So, in this study like once again the potential functions are defined in the same way; however, the positive and negative anomaly events are handled separately in the same framework as earlier. And here the parameters are chosen in is like, so that they can make a trade - off between the spatio-temporal extents of the events and the intensity of the identified events. That is like we can go for like intense anomaly events which might have had a smaller spatio-temporal spread or we may go for larger I mean less intense anomaly events which may have a larger spatio-temporal spread and so on.

So, and also these potential functions they have the ability to filter out the either the mild or the localized events and identify the intense and extensive ones. And like in this case also we find that some events are they are single-year and the others you span many years. So, like these are just some examples of the different anomaly events that were identified by this paper so, there were some positive anomaly events. So, like for example, if you consider the years 1994 and 2001.

So, these yellow regions these regions were found to like you can see these are somewhat extended regions covering several grid points, these had some kind of positive anomaly events of rainfall. Similarly, in the years 2000 and 2002 both of which were deficient rainfall years or deficient monsoon years. So, we can see that these were the regions which had large negative anomaly events of rainfall that is they received much less rainfall than usual. Similarly, like in this study they identified various events which lasted several years and so on.

(Refer Slide Time: 18:40)



Now, we come to another topic another related topic, here the aim is to use deep learning for extreme event identification. The aim here is to identify like certain extreme weather events like weather fronts, tropical or extra-tropical cyclones etcetera from high resolution simulations by models or like it can also be from observations, but usually the observations do not give are not that high resolution yet especially in the historical period where we did not have such so much good say powerful satellites as we now have. So, we have to depend on the model simulations for the high to get the high resolutions values of many variables.

So, these extreme weather events they usually manifest in several climatic variables as a spatio-temporal function. So, when there is a cyclone it is unlikely to be manifested in only one type of like variable like precipitation or something like that, but it can manifest in multiple things like u wind speed. I mean by u wind I mean meridional wind, the zonal winds etcetera as well as may be like the they may be associated with low pressure etcetera.

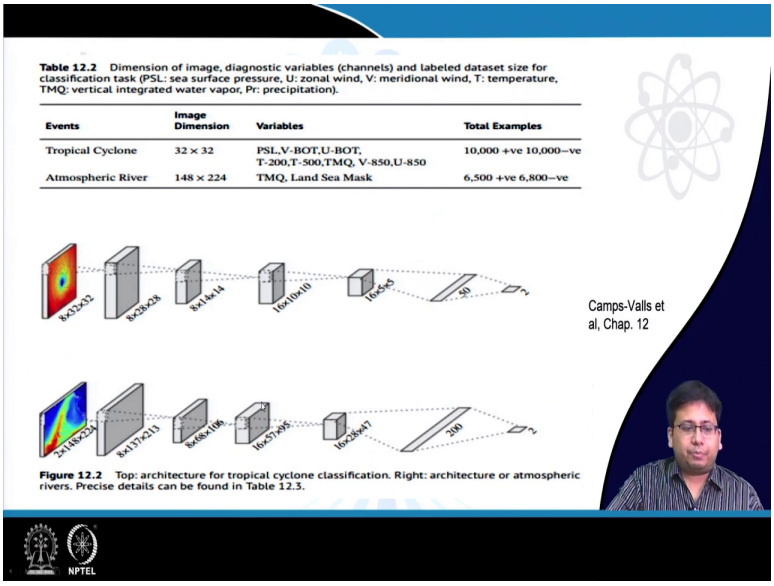
So, these things manifest themselves in various variables. So, we have to consider all these variables into account while identifying these extreme events. So, the like what the way it used to be done earlier was like this. So, we have the spatial maps of different variables of interest that was the input. So, like we actually used to carry out the domain experts used to extract the various features locally by calculate using some mathematical functions or whatever.

And then there were some heuristic algorithms say something based on typically based on thresholds etcetera, also based on some kind of computer vision, based tracking or something like that to I will and then once they have done all these things then they could do some simple classification to identify whether there is a cyclone there or not.

So, that used to be the approach earlier, but then again that requires the intervention of the domain experts, it requires setting of thresholds and we have already discussed that threshold is always a dicey issue. So, what is now done is that, the inputs I mean the input maps of the different variables they are all provided as something like as inputs to like a deep learning or a model which first of all it treats all those spatio-temporal data of the different inputs input variables as like as spatial maps or images.

So, just like in the previous lecture we were talking about representing the spatial maps as images and using the SR-CNN kind of models. So, in this case also we treat these spatial maps as images and apply some kind of neural network on them and then the these neural networks are often used for say some tasks like image classification that is given an image it predicts is it a cats image or a dogs image. So, that same like that is basically something like a classification image classification task, in this case also they do image classification where they classify the spatial maps and just say is there a cyclone or is it there or not. So, that is the general approach.

(Refer Slide Time: 22:18)



So, like here like in this paper like they have talked about the identification of two extreme events tropical cyclones and atmospheric river. And there for this purpose they as I said different extreme events they manifest themselves in several variables. So, here these are the different variables they have considered that is the sea level the sea surface pressure, the zonal wind, meridional wind, temperature, the vertical integrated water vapor, precipitation etcetera.

So, like the input is these maps the of the spatial maps of all these variables. So, just like a normal image has three channels the R, G and B channel, here we can say that each of these variables that we have considered each of them is one channel. So, the in this kind of a map we there are there might be 8 channels or in this case there might be 2 channels etcetera.

So, we have an image like this. So, that which is the spatial map and then that as it happens in the case of a image classification; the image is passed through a series of convolution and pooling layers in each layer there might be lots of convolution filters stacked parallelly and each of those filters may have many channels also. So, like with each convolution operation we get progressively smaller and smaller images they are white and I mean apart from convolution there are pooling also which is like make the which as I said earlier in the earlier lectures.

They throw away the less relevant pixels and store only the more relevant ones and finally, they reach the scalar output which can take only two values that is tropical cyclone or no tropical cyclone.

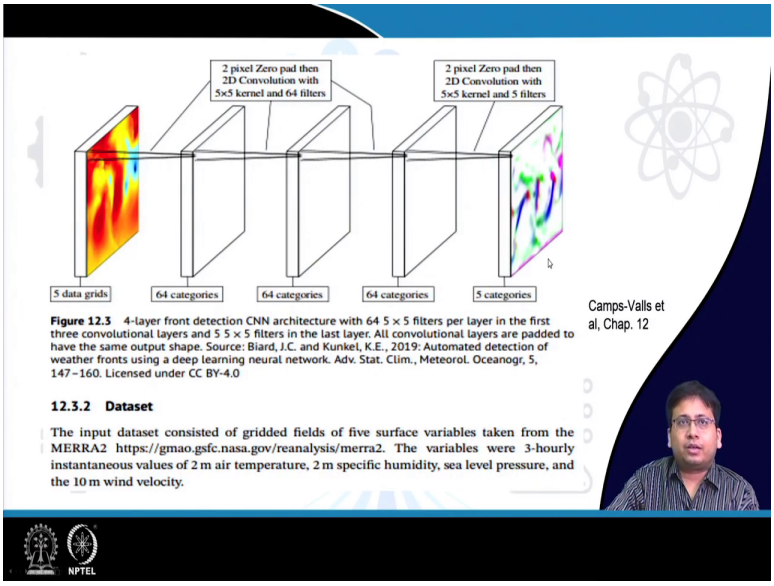
Similarly, in like for the atmospheric river also we can do this. So, like tropical cyclone is something is a more localized event that is of the when we are considering the map of a particular variable all over the world. I mean the cyclone is not something that takes place over a wide region, it will take place over only a small region or we so we should really be focusing on small image patches.

So, that is why they are focusing on this $32X32$ images, which is like we can say it is a small region. I will study the maps of the spatial of these 8 different variables over a small region. Let us say over some part of bay of the Northern Bay of Bengal or not and then classify that whether there is a cyclone there or not, but atmospheric river this is like its a widespread phenomena, it is

something that manifests itself over a much wider scale. So, that is why they have chosen a much larger dimension $148X224$.

But the basic thing remains the same, that same image is passed through a series of operations of convolution and pooling and ultimately we reach the this output this scalar output which just says atmospheric river is there or no atmospheric river is there.

(Refer Slide Time: 25:43)



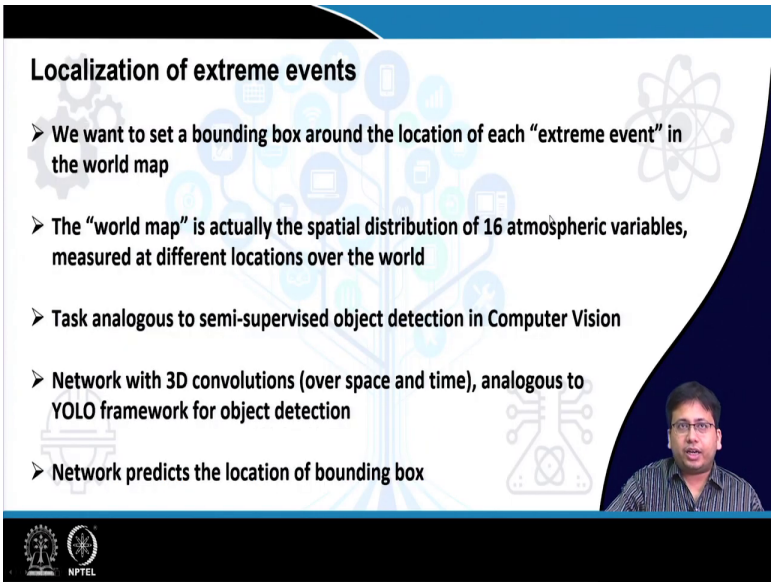
So, this like this is what the some of these neural networks actually look like. So, suppose I want to do the like the 5 categories, let us say I want to like giving us a large spatial map over a large region I want to like identify 5 different kinds of extreme events. So, there are 5 like at every pixel I want to classify it as one of 5 categories, one category is of course, that no weather, no extreme event, another category might be tropical cyclone, another category might be atmospheric river etcetera.

So, like in the. So, the input is at every pixel or every grid point I have the measurements of the different variables. The output is at again like a map where at every pixel I am like indicating what kind of extreme event is there or no extreme event is there. So, in this kind of a classification problem also we do it with the help of this like convolution neural network. So, again we have a series of convolution layers like this in which 2D convolution takes place with

kernels and each and each layer there are large number of filters which are like placed with each other.

So, the and so, that this is like the categorize I mean like the mapping problem that is we have the world the map of different variables and the target is to like at every pixel identify like which event is taking place.

(Refer Slide Time: 27:24)



An alternative version of this problem is the localization of extreme events where we want to set a bounding box around the location of each "extreme event." So, the like once again we take the "world map" which is actually the precipitate the spatial distribution of 16 atmospheric variables, measured at different locations all over the world.

And so, the problem with these kind of that is approach is that typically computer vision or this deep learning based computer vision algorithms, they require a very large amount of labeled data. And so, in this case the data is coming from this simulated world maps, but then the problem is the labeling that is who will actually sit and do the labelings.

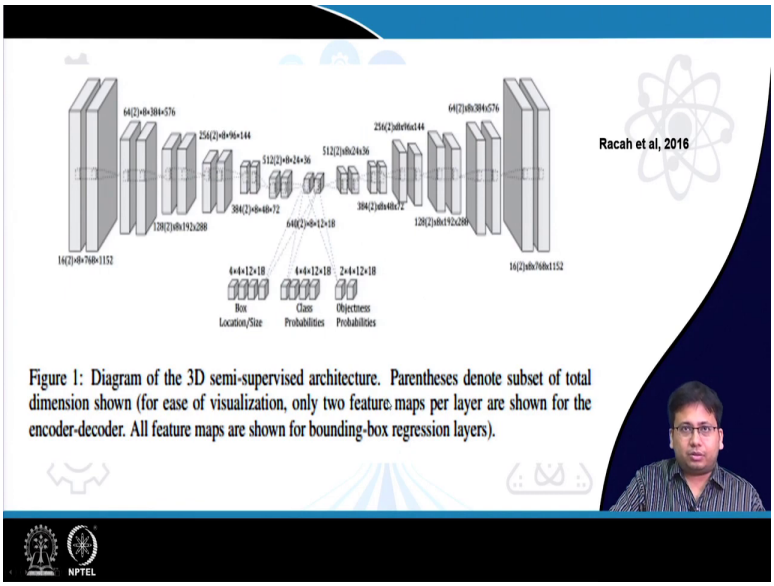
So, like typically what happens is only some images are labeled, in the other images even for training they are they may might be unlabeled. So, we have to formulate it as a semi supervised learning task in which some of those unlabeled images they are also utilized for the training

purpose. And so, like what we are trying to do here is that given a the spatial map of a particular variable we are trying to like find the bounding boxes of the of the phenomena which we are interested in tropical cyclones etcetera.

So, it is something like an object detection problem. So, for in the computer vision object detection is often done using a framework called YOLO which is based on again convolution neural network. So, in this case also we use something analogous to YOLO and in like we use in this case 3D convolution that like 3D as in whether we consider time also that is we consider convolution over time also, because these things especially because like we have data over high spatial temporal resolution also.

So, like even maybe we have 3 hourly data etcetera. So, instead of focusing on only one time point we can actually focus on say the several time points to identify whether the quantity of interest that is present or not.

(Refer Slide Time: 29:36)

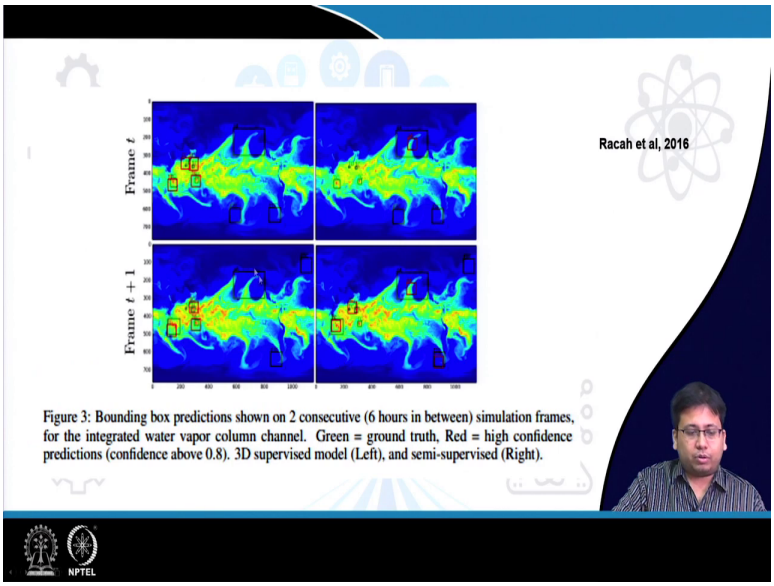


So, this is what the network structure looks like. So, the it is basically something in the something like an auto encoder. So, the like it is an auto encoder only the task of the auto encoder is to take the spatial map of the variables as input and reconstructed. But in the process of reconstruction it reduces the like input to a like a to a small dimension to a very low

dimension that we have earlier also discussed the how auto encoder does the dimensional reduction.

So, once you have this low dimensional or the encoded representation it is like we can use some kind of a classification and or we basically we can use another neural network to identify the exact box the bounding box of the particular event and also the class probabilities that is I mean the different classes are the different types of anomalies. So, to like every small box or every small region we associate with it the probability of whether it contains some kind of an anomaly event or not. So, that is how this whole thing works.

(Refer Slide Time: 30:56)



And this is the possible output. So, like you can see some red or boxes here, these red boxes they indicate the high confidence prediction that is these are the locations where they are quite confident that there is a tropical cyclone of or whatever they are interested in and there are these green boxes like you can see the these green boxes these are the ground truth locations of the of the tropical cyclones.

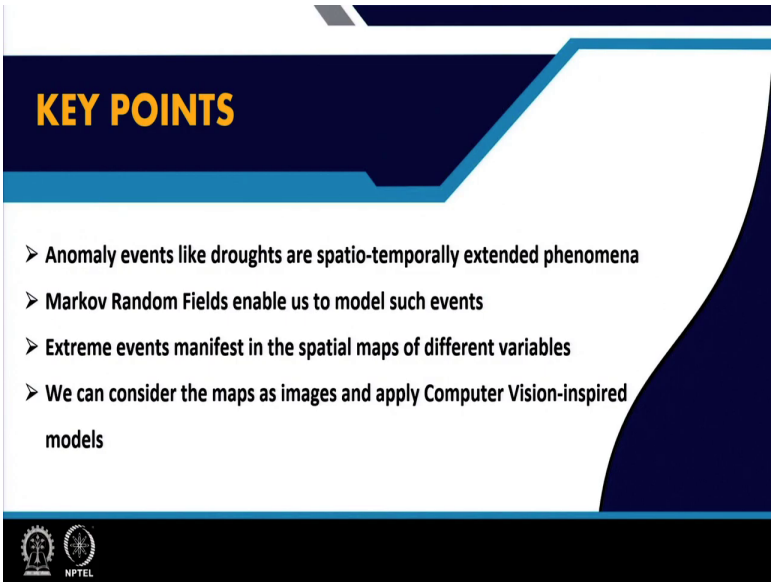
So, you can see in many cases there is some reasonable overlap between the red boxes or the green boxes and there are also some situations where the red box is only a like a small part inside the green box or where the green box is actually a small part of the red box etcetera. And so, like

and as you can also see that this is done continuously over time and so that we can do the 3D convolution and the boxes that are identified at any given time they help in identifying the boxes at the next point of time also.

(Refer Slide Time: 32:00)

So, these are the set of papers which we discussed the for the first two are related to the anomaly detection using the anomaly event detection using Markov random fields and the next are the ones where we used deep learning to use identify the extreme events.

(Refer Slide Time: 32:17)



So, the key points to be taken away from this lecture are that the anomaly events like droughts are spatio-temporally extended; Markov random fields enable us to model such events. The extreme events they manifest in the spatial maps of several variables and we can consider these maps as images and apply computer vision inspired models on them.

So, that brings us to the end of this lecture. In the coming lectures also we will consider other ways in which machine learning can be used to solve various or to find new insights related to the earth sciences.

So, till then goodbye.