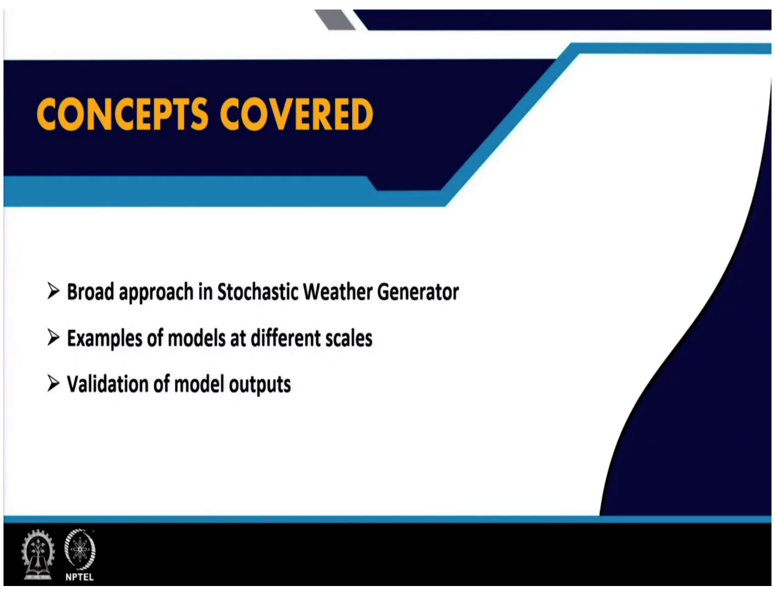
Machine Learning for Earth System Sciences
Prof. Adway Mitra
Department of Computer Science and Engineering
Centre of Excellence in Artificial Intelligence
Indian Institute of Technology, Kharagpur

Module - 05
Machine Learning for Earth System Modelling
Lecture - 35
Stochastic Weather Generator

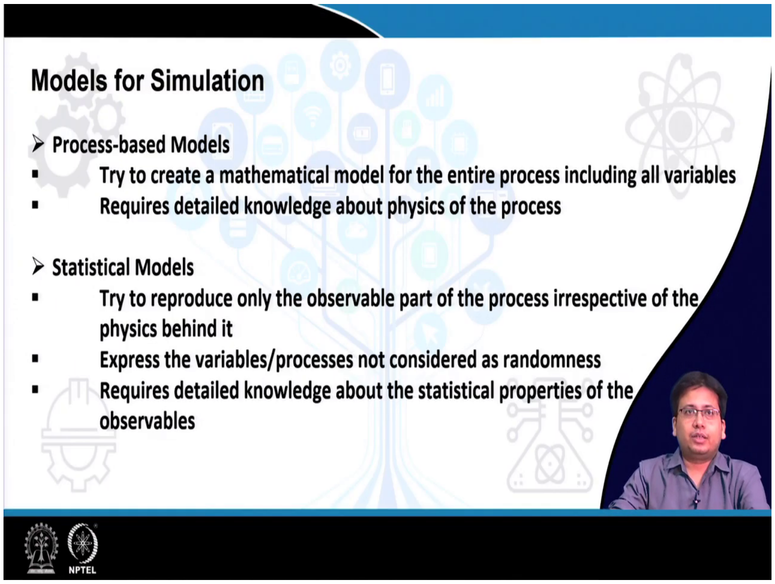
Hello everyone. Welcome to lecture 35 of this course on Machine Learning for Earth System Science. We are currently in module 5, our last module where we are discussing Machine Learning for Earth System Modelling. The topic of this lecture is going to be Stochastic Weather Generator.

(Refer Slide Time: 00:42)



Now, in this lecture, we will discuss the broad concepts of stochastic weather generators, what they are, how they work etcetera, like we will see some examples of such models at different scales and applications. And we will also see like the outputs they provide by their simulations, how we can validate them.

(Refer Slide Time: 01:03)



So, first of all, just an idea about what this thing is all about. So, like we have earlier also discussed process-based models where we try to create the mathematical model of a particular earth system process which includes all the variables. And such process-based models require the detailed knowledge of the physics of the process including the different variables, how they are related to each other, how they influence each other and so on. So, typically the process-based models are based on a set of equations which are often known as the governing equations of the model.

Now, the opposite of the process-based models are the statistical models where we try to reproduce only the observable parts of the process irrespective of the physics behind it. That means, like we do not really care about those variables which are not observed or which are not relevant.

Like, we can approximate those variables by introducing some like conceptual variables which are, which do not actually correspond to any physical quantity. Like for example, the weather state and things like that which we may have discussed earlier. Typically, those are latent variables.

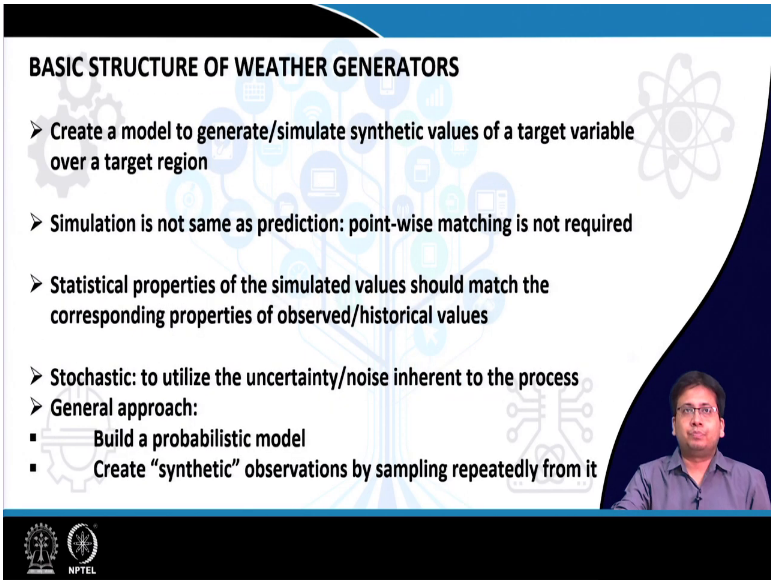
And then those variables and processes which are not considered as part of the system nor are like provided as external covariates, they are considered as randomness. That is we like; so, statistical models are typically like stochastic models based on probability distributions. And this probability distributions basically they account for the uncertainty of the system.

Why is there uncertainty in the system? That is precisely because various small scale processes and so on, they like, they are not explicitly modeled in the statistical these statistical models. They, like in the, in case of the process-based models like it is expected that all the processes like I mean all the sub-processes irrespective of how small scale they may be, they should be included. And that is why these process-based models are ideally should not contain any randomness because they should have accounted for all the sources of random possible sources of uncertainty in the model itself.

So, these process-based models are deterministic models. In reality of course, they may often not be able to handle all the, like all the sub-processes. We have, earlier in some lecture we have discussed about unresolved processes and so on. So, those are like a typically a weakness of the process-based models. In case of statistical models, they are simply expressed as like as some kind of uncertainty which is quantified by probability distributions.

And unlike the governing equations of the process-based models, statistical models typically have lots of probability distributions. Conditional distributions which indicate how one variable or what values one variable may take provided the another variable takes some other values and so on.

(Refer Slide Time: 04:11)



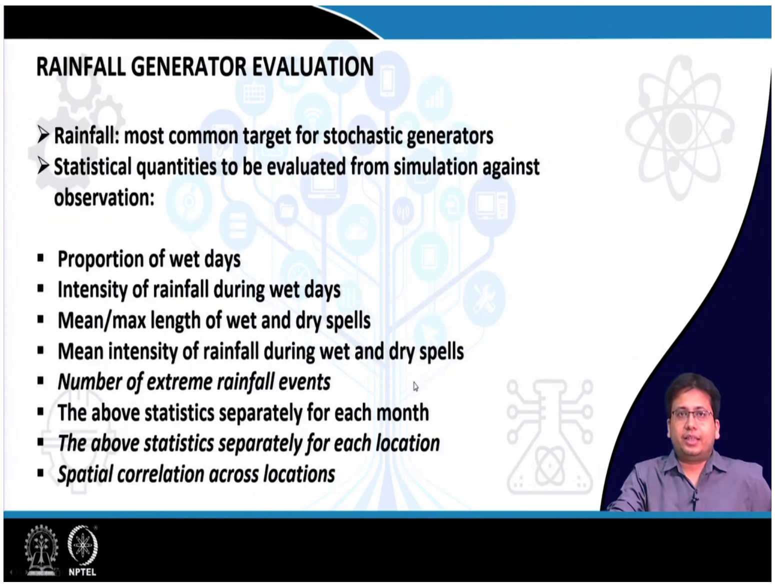
So, now, so like, so these statistical models we can say they are some kind of lightweight process or earth system models. While an earth system model is supposed to be a like a large scale model for a which like typically covers a large area and handles the multiple variables and so on. A statistical model is supposed to be a small scale lightweight model that may be focused on only one particular region and may be focusing on only 1 or 2 variables of interest.

So, the structure of the now this weather generator is such a statistical model. It is a very like an, it is a this weather generators independently of have been studied in disciplines like hydrology and civil engineering, since maybe the late 80s and so on. Of course, now they have become much more developed than earlier. But they can be considered as a small scale earth system process models.

So, the general approach is as follows. Create a model to generate the synthetic values of a target variable over a target region. And this simulation is not the same as prediction. That is we are not trying to say that at a particular time t, the value of the variable at a particular location s will be this much. I am not aiming to make forecasts of anything like that. I will just generate synthetic data and it is really the statistical properties of this like synthetic or simulated values which should match the corresponding properties from the observed or historical values.

Now, the like, as I have already said the stochasticity is added to like capture the uncertainty which is in the process; and so and the so like it will basically be a probabilistic model. And the observation, and the simulated values which we are talking about will be obtained by sampling repeatedly from that probabilistic model. That is we will make a series or sequence of values sampled from it and that should give us the like one particular realization of the simulation.

(Refer Slide Time: 06:21)



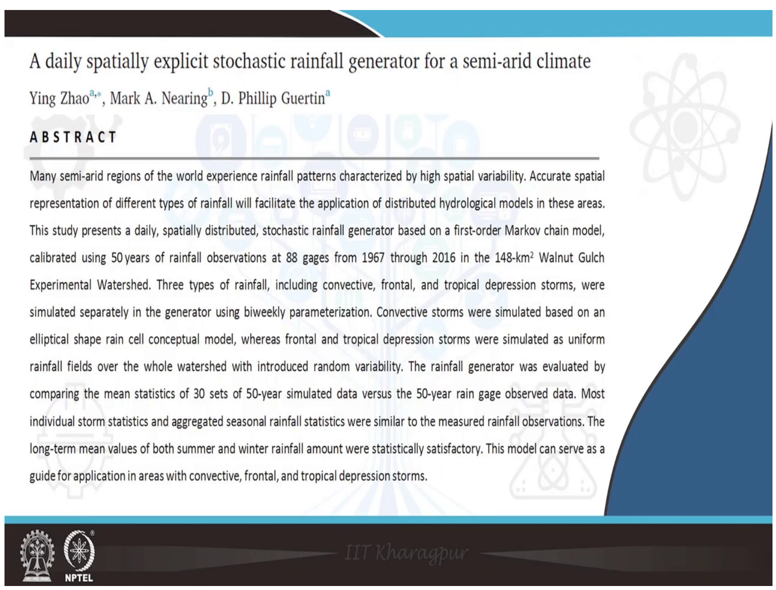
Now, among stochastic weather generators, the variables that are most commonly targeted are rainfall and temperature. So, like there is a whole bunch of papers called the stochastic rainfall generator which is a subset of stochastic weather generation. So, like, when we, as I already said we really need to compare the statistical properties of the simulation with what has already been observed historically.

So, in case of rainfall simulation like this, the properties that we study are as follows. It might be the proportion of wet days in a season. Wet days meaning those days where there was nonzero rainfall or there was rainfall above a particular threshold. Or then what is the average intensity of rainfall during such wet days? Then, these wet days or dry days they are usually they are not isolated, but they come in spells, that is maybe 3 successive rainfall rainy days at a stretch or maybe another 5 or 10 successive dry days at a stretch and so on.

So, these are spells. So, what are the minimum maximum and average value lengths of such spells? And what is the I mean intensity of rainfall during say wet spells? And then how many extreme rainfall events happened? We have already discussed the concept of extreme in earlier lectures. So, how many such extreme events happened. Then, so these are like we can say these are gross statistics.

We can; like these things we can define like calculate for one particular location over the region in, over which we are simulating or we can like calculate all these things for a specific months and besides like we can also calculate things like spatial correlation of the variable across the different locations. So, earlier we have talked about like variograms and things like that. So, like the is the simulated variogram similar to the observed variogram. So, these are questions which we are typically interested in when we are like evaluating the output of a stochastic rainfall generator.

(Refer Slide Time: 08:34)



So, let us consider a paper, recent paper which appeared in related to this topic. So, here a daily spatially explicit stochastic rainfall generator for a semi-arid climate. So, what is spatially explicit? Let us try to understand that. Many semi-arid regions of the world experience rainfall patterns characterized by high variability, high spatial variability.

Accurate spatial representation of different types of rainfall will facilitate the application of distributed hydrological models in these areas. This study presents our daily, spatially distributed, stochastic rainfall generator based on first-order Markov chain model, calibrated using 50 years of rainfall observations at 88 gauges from 1967 to 2016 in the 148 square kilometer region which is the experimental region. This is somewhere in the US.

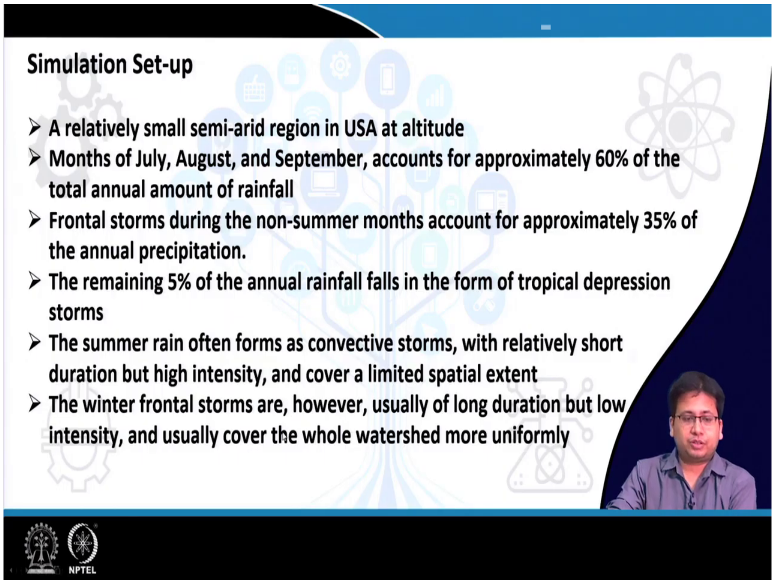
Three types of rainfall, including convective, frontal, and tropical depression storms, were simulated separately in the generator using biweekly parameterization. So, we know that rainfall can happen in different ways. Like, there can of course, be the normal convective rainfall which is brought about let us say during the monsoon season and so on. And then there can be storms, tropical storms, cyclones or low pressure, depressions and so on.

So, they like different types of rainfall have different kinds of characteristics. And during the summer months we have these rainfalls causing by and well, as a result of the Nor'westers and things like that. So, they will like. Now, each type of rainfall may have different characteristics, some might be spatially more focused, some may be spread over the entire region, some may have it like happen only in certain hours of the day, while the others can happen anytime and so on. So, in this work, they are trying to handle all the 3 categories separately and simulate their statistics accordingly.

The convective storms were simulated using an elliptical shape rain cell conceptual model, whereas frontal and tropical depression storms were simulated as uniform random fields over the whole waters shell with introduced random variability. The rainfall generator was evaluated by comparing the mean statistics of 30 sets of 50 year simulated data versus 50 year gauge observed data.

Most individual storm statistics and aggregated seasonal rainfall statistics were similar to the measured rainfall observations. The long-term mean values of both summer and winter rainfall amount where statistically satisfactory. This model can serve as a guide for application in areas with convective, frontal, and tropical depression storms.

(Refer Slide Time: 11:45)

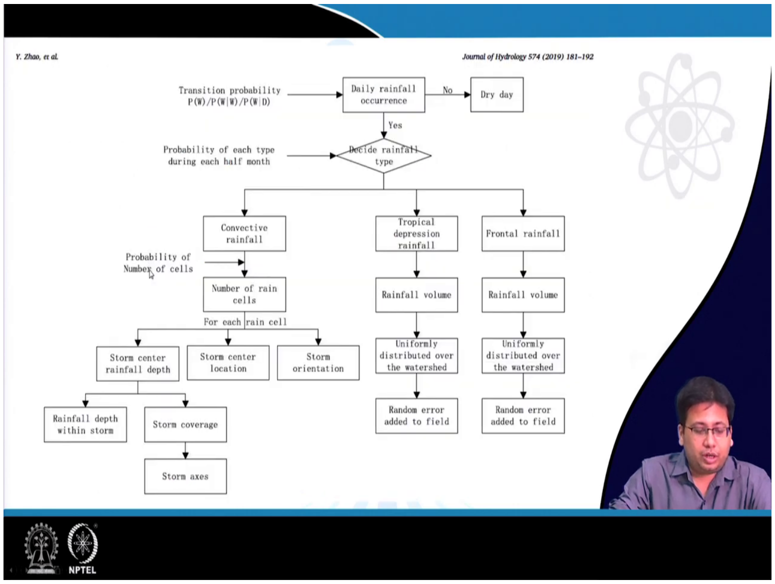


So, here they have they the setup is as follows. So, they are targeting a relatively small semi-arid region of US at a high altitude. So, the characteristics of this region is such that the summer months of July, August, September, they typically account for approximately 60 percent of the total amount of rainfall. Just like in India, the most of the monsoon rainfall, like most of the places receive 80 percent of their annual rainfall from monsoon which is from June to September. Here also it is more or less the same case, except that it is not a monsoon because it is not a tropical region.

Frontal storms during the non-summer months account for approximately 35 percent of the annual precipitation. And the remaining 5 percent of the annual rainfall happens in the form of tropical depression storms. Now, the summer rain often like forms as convective storms, with relatively short duration, but high intensity, and cover a limited spatial extent. The winter frontal storms are usually of a long duration, but low intensity and usually cover the whole watershed more uniformly, ok.

So, these are the main characteristics of the different types of rainfall events. Note that here each type of event, we have actually characterized which in which months they are more likely to happen, and when they happen how they are spread spatially, ok.

(Refer Slide Time: 13:12)



So, based on that this is the like the rough structure of the model which they will build. So, first of all there is a daily rainfall occurrence, I mean is it going to rain or not, is it a dry day or not. So, and so this is actually considered as some kind of a first order process, like because we know that if yesterday was a dry day probably today will also be a dry day, if yesterday was a wet day probably today will also be a wet day.

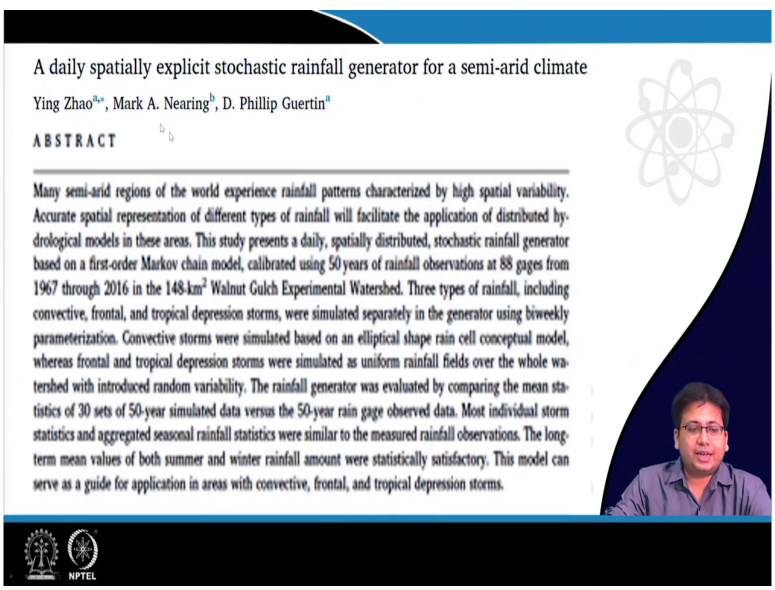
And apart from a few wet to dry, your dry to wet transitions can happen with certainly and probability. So, based on that, first of all it is like the model decides if any given day is going to be wet or dry over the entire region. Suppose, it is a dry day, in that case of course, there is nothing to simulate. If there is a wet day, then they will decide what kind of rainfall is going to happen. So, out of the 3 kinds discussed earlier.

Now, we have also seen that in the different months each rainfall, each type of rainfall has a certain probability. So, like which; so, if it is going to be this daily rainfall occurrence; like if it is considered to be a wet day, the next thing we will see is like which type of rainfall will happen and that will depend on the month. It will of course, also depend on the previous day I mean if previous day, like one kind of event happened, rainfall event happened probably the same thing will continue today also. So, that is that plays a role. Besides what month it is also plays a role.

Now, suppose we have decided or suppose, so this is decided by a sampling. So, suppose we have decided that it is going to be a frontal rainfall, so once if it is a frontal rainfall, then the rainfall volume will be simulated by drawing from some suitable probability distribution which is characteristic of frontal rainfall events. And this rainfall volume, we will distribute over the entire region. That is equal all locations will get roughly equal amount of rainfall with some amount of error which will be added randomly.

Similar thing will happen if it is decided that it is a tropical depression. Then, also we will equally spread all over the region. But if it is a convective rainfall which we know is a now we read earlier, is a localized process. So, in that case, it will have to spatially like or explicitly we will have to specify the spatial location, where exactly it is going to happen. That is why they call it as a spatially explicit model.

(Refer Slide Time: 15:49)



So, it will decide how many that is, the area that it is going to cover that is the number of rain cells. So, like, is it going to cover only a small number of grid cells or will it cover a large number of grid cell and so on. And then, for each of those cells which in which the rainfall will happen, like different further characteristics of the storm like where exactly its center will be, what will be its depth what will be its orientation etcetera. All these things will be drawn from appropriate distributions. And then the accordingly the volume of rainfall will be simulated.

(Refer Slide Time: 16:33)

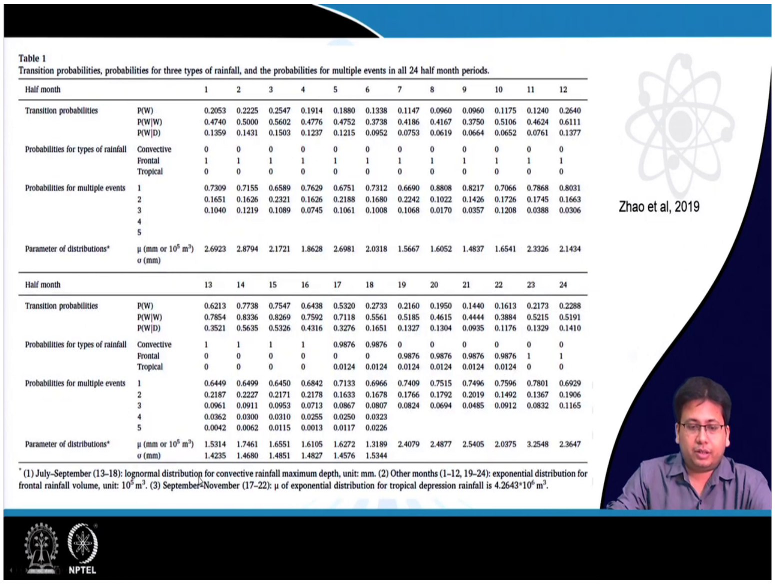


Table 1
Transition probabilities, probabilities for three types of rainfall, and the probabilities for multiple events in all 24 half month periods.

Half month		1	2	3	4	5	6	7	8	9	10	11	12
Transition probabilities	P(W)	0.2053	0.2225	0.2547	0.1914	0.1880	0.1338	0.1147	0.0960	0.0960	0.1175	0.1240	0.2640
	P(W|W)	0.4740	0.5000	0.5602	0.4776	0.4752	0.3738	0.4186	0.4167	0.3750	0.5106	0.4624	0.6111
	P(W|D)	0.1359	0.1431	0.1503	0.1237	0.1215	0.0952	0.0753	0.0619	0.0664	0.0652	0.0761	0.1377
Probabilities for types of rainfall	Convective	0	0	0	0	0	0	0	0	0	0	0	0
	Frontal	1	1	1	1	1	1	1	1	1	1	1	1
	Tropical	0	0	0	0	0	0	0	0	0	0	0	0
Probabilities for multiple events	1	0.7309	0.7155	0.6589	0.7629	0.6751	0.7312	0.6690	0.8808	0.8217	0.7066	0.7868	0.8031
	2	0.1651	0.1626	0.2321	0.1626	0.2188	0.1680	0.2242	0.1022	0.1426	0.1726	0.1745	0.1663
	3	0.1040	0.1219	0.1089	0.0745	0.1061	0.1008	0.1068	0.0170	0.0357	0.1208	0.0388	0.0306
	4												
	5												
Parameter of distributions*	μ (mm or 10^5 m^3)	2.6923	2.8794	2.1721	1.8628	2.6981	2.0318	1.5667	1.6052	1.4837	1.6541	2.3326	2.1434
	σ (mm)												

Half month		13	14	15	16	17	18	19	20	21	22	23	24
Transition probabilities	P(W)	0.6213	0.7738	0.7547	0.6438	0.5320	0.2733	0.2160	0.1950	0.1440	0.1613	0.2173	0.2288
	P(W|W)	0.7854	0.8336	0.8269	0.7592	0.7118	0.5561	0.5185	0.4615	0.4444	0.3884	0.5215	0.5191
	P(W|D)	0.3521	0.5635	0.5326	0.4316	0.3276	0.1651	0.1327	0.1304	0.0935	0.1176	0.1329	0.1410
Probabilities for types of rainfall	Convective	1	1	1	1	0.9876	0.9876	0	0	0	0	0	0
	Frontal	0	0	0	0	0	0	0.9876	0.9876	0.9876	0.9876	1	1
	Tropical	0	0	0	0	0.0124	0.0124	0.0124	0.0124	0.0124	0.0124	0	0
Probabilities for multiple events	1	0.6449	0.6499	0.6450	0.6842	0.7133	0.6966	0.7409	0.7515	0.7496	0.7596	0.7801	0.6929
	2	0.2187	0.2227	0.2171	0.2178	0.1633	0.1678	0.1766	0.1792	0.2019	0.1492	0.1367	0.1906
	3	0.0961	0.0911	0.0953	0.0713	0.0867	0.0807	0.0824	0.0694	0.0485	0.0912	0.0832	0.1165
	4	0.0362	0.0300	0.0310	0.0255	0.0250	0.0323						
	5	0.0042	0.0062	0.0115	0.0013	0.0117	0.0226						
Parameter of distributions*	μ (mm or 10^5 m^3)	1.5314	1.7461	1.6551	1.6105	1.6272	1.3189	2.4079	2.4877	2.5405	2.0375	3.2548	2.3647
	σ (mm)	1.4235	1.4680	1.4851	1.4827	1.4576	1.5344						

* (1) July–September (13–18): lognormal distribution for convective rainfall maximum depth, unit: mm. (2) Other months (1–12, 19–24): exponential distribution for frontal rainfall volume, unit: 10^5 m^3. (3) September–November (17–22): μ of exponential distribution for tropical depression rainfall is $4.2643*10^6$ m^3.

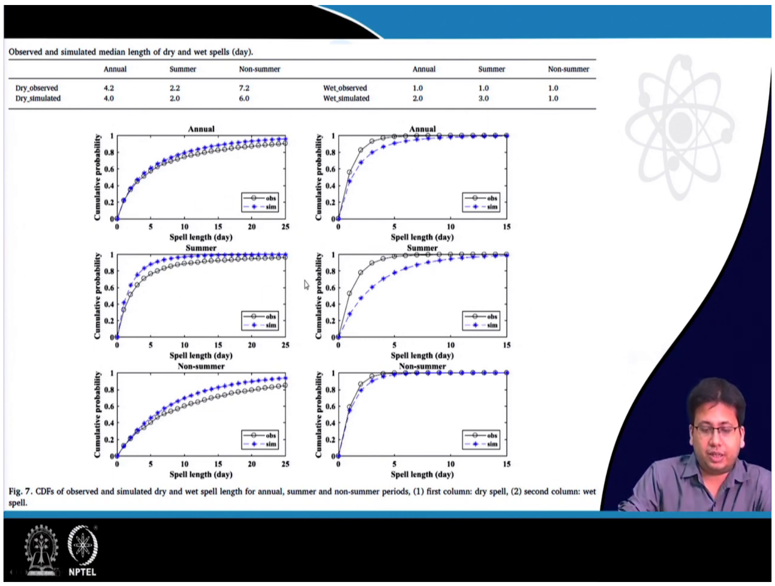
So, like these are typical parameter values of each of those probabilities which are estimated based on the observations. So, most of these probabilities are actually nothing but relative frequencies, like which are obtained from the observation data. So, they as they mentioned, they have about 50 years of data, 50 years of daily scale data which is quite a lot of data.

And then accordingly, so for the different months, they have divided like, they have like estimated the probabilities of all these things. So, as you can see the transitional probability from that is from wet to wet, wet to dry etcetera; the probabilities of the 3 different types of rainfall and so on. So, as you can see the entire year has been divided into 24 half months period.

Now, as you can see in the first 6 months that is from January to May, if rainfall happens it will surely be frontal, frontal storms. But if you consider July and August, then the rainfall is almost surely going to be of the convective type. And if you consider September, mostly it will be convective, but there is a small chance that it might be tropical also. As the; but if you consider the following months, it is most likely going to be a frontal even though tropical there is a small chance of and so on and so forth.

Then, the what is the probabilities of multiple events happening on the same day? So, that is also estimated and then the other various parameters of the distribution. So, like these amount of rainfall volume and so on. So, these are typically handled using some kind of Gaussian or maybe non-Gaussian distributions. So, they will have certain parameters of the model, the μ, σ etcetera. So, those are also estimated from the observations.

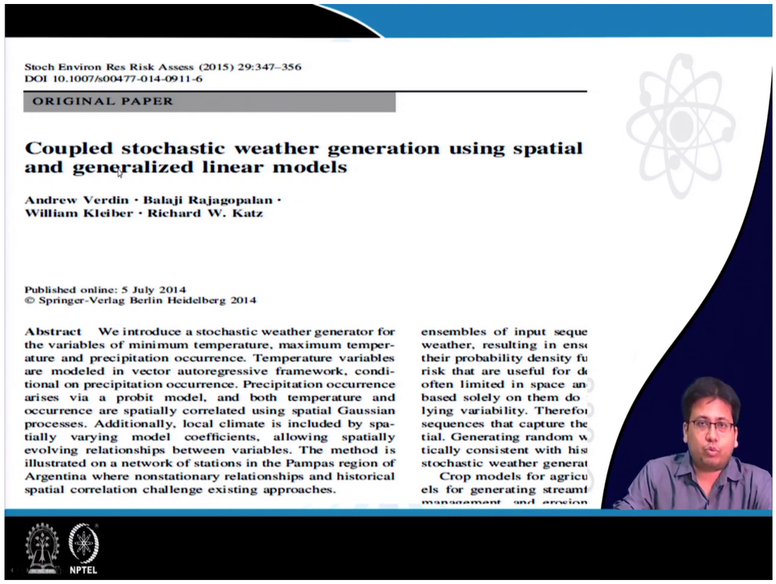
(Refer Slide Time: 18:31)



Now, if we go to the simulation results. So, like as I said, the data is obtained by drawing samples from all these like from this kind of a model. So, 50 years of observed data they have. So, they keep on drawing the data for 50, another 50 artificial years. So, for every day and every location, they will simulate the amount of rainfall using this model. And then, they will calculate the various statistics of that model with the corresponding statistics from the observations.

So, the like the spells of the wet days, the spells of the like all over the year, then only in the summer, in the non-summer months, what is the typical duration of a wet spell or what is the typical duration of a dry spell. So, like as you can see they have; so, the this is the what they have obtained from the observations and this is what they have obtained from simulations. So, like as you can see the like each particular length of a kind of spell, it has a probability. So, they have basically plotted the CDF of that.

(Refer Slide Time: 19:45)


Stoch Environ Res Risk Assess (2015) 29:347–356
DOI 10.1007/s00477-014-0911-6

ORIGINAL PAPER

Coupled stochastic weather generation using spatial and generalized linear models

Andrew Verdin · Balaji Rajagopalan · William Kleiber · Richard W. Katz

Published online: 5 July 2014
© Springer-Verlag Berlin Heidelberg 2014

Abstract We introduce a stochastic weather generator for the variables of minimum temperature, maximum temperature and precipitation occurrence. Temperature variables are modeled in vector autoregressive framework, conditional on precipitation occurrence. Precipitation occurrence arises via a probit model, and both temperature and occurrence are spatially correlated using spatial Gaussian processes. Additionally, local climate is included by spatially varying model coefficients, allowing spatially evolving relationships between variables. The method is illustrated on a network of stations in the Pampas region of Argentina where nonstationary relationships and historical spatial correlation challenge existing approaches.

Similar, along similar lines we discussed another paper, so here coupled stochastic weather generator using spatial and generalized linear models. So, just like, in the previous case, the buzzword was spatially explicit. That is how it was difference, this model was differentiated from the other models. So, like by the way for the evaluation of the model, so this spell length is only one of the criteria by which they have compared the simulations with the actual results. Apart from that there are bunch of other parameters statistical quantities also which I already discussed.

So, they all those things also have been compared against the observations. So, you are strongly encouraged to go back to these papers and actually read it. In the next paper, here it is called coupled model because it like; here it is not only about rainfall. It has two variables which are dependent on each other. So, the coupling of these two variables, namely temperature and rainfall is actually captured or attempted to be captured by some kind of a probabilistic model.

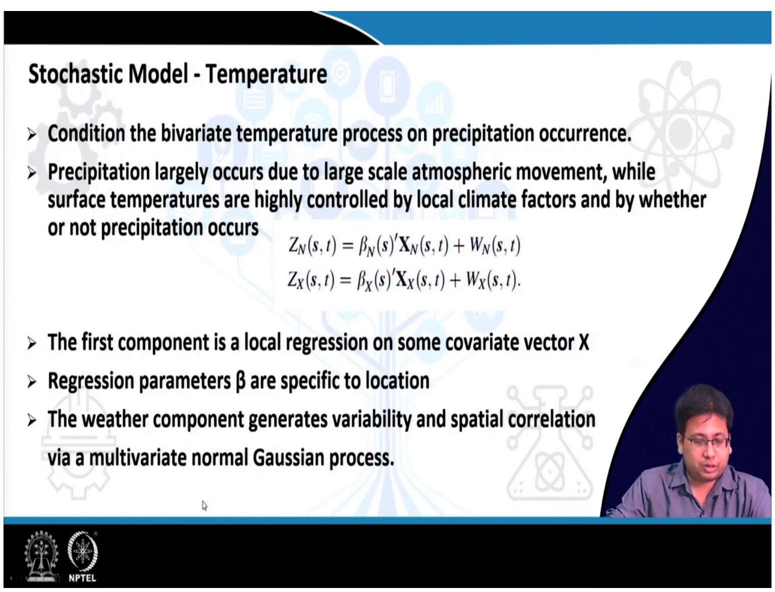
So, we introduced a stochastic weather generator for the variables of minimum temperature, maximum temperature and precipitation occurrence. Temperature variables are modeled in vector auto-regressive framework, conditional on precipitation occurrence. So, auto-regression we have discussed earlier. So, but it is not just auto regression, that is it depends not only on its own past values, but also upon precipitation occurrence.

That is like if there is precipitation then today, then tomorrow's temperature will depend on today's temperature in a particular way. But if there is no rainfall today then tomorrow's temperature will depend on today's temperature in a slightly different way. Precipitation occurrence arises via a probit model, and both temperature and occurrences are spatially correlated using spatial Gaussian processes.

So, we have discussed these Gaussian processes earlier also. And we have also particularly discussed how Gaussian process is required to maintain the spatial correlations between the different for the different variables. That is if so much rainfall happens in location $s1$, how much rainfall will happen in location $s2$, these kinds of things.

So, we have seen that Gaussian processes are a good way of or a or a it is a good model which is able to specify these kinds of correlations through the covariance function. So, that is what they do here. Additionally, local climate is included by spatially varying model coefficients allowing spatially evolving relationships between variables. The method is illustrated on a network of station in the Pampas region of Argentina where non-stationary relationships and historical spatial correlation challenge existing approaches.

(Refer Slide Time: 22:50)



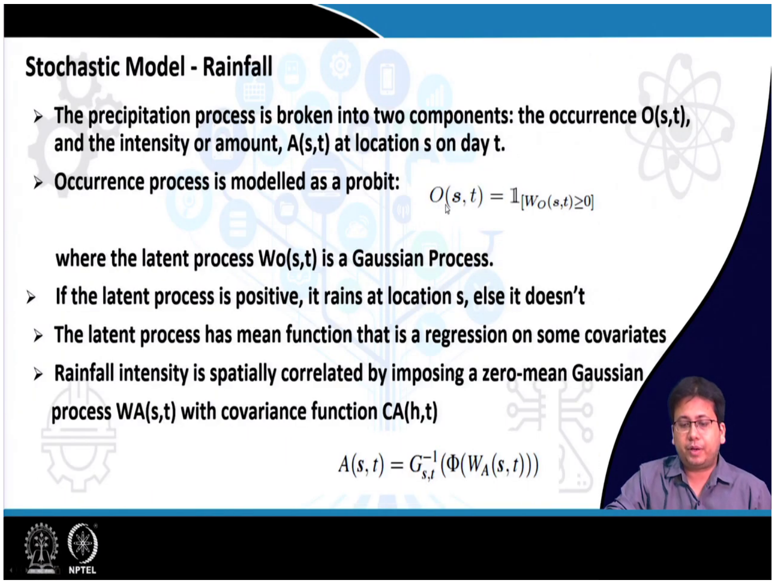
So, as I said here are two different variables to be simulated here temperature, in fact, maximum and minimum temperature and rainfall. So, like the bivariate temperature process, bivariate because they are there are two things maximum and minimum, that temperature process is conditioned on the precipitation occurrence. Now, the precipitation occurs largely due to large scale atmospheric movement, while the surface temperature are highly controlled by local factors and also by whether or not precipitation happens.

So, for the time being let us keep precipitate the, let us just assume that the precipitation is a covariate as far as the temperature is concerned. So, at a given location on a given day the maximum temperature indicated by Z_X and the minimum temperature indicated by Z_N, they follow like certain characteristics. So, X_N, X_X these are basically the covariates which influence the temperature. This also includes the precipitation which is being simulated.

So, these β_X, β_N, these are the correlation coefficients as sorry; I mean the regression coefficients. And so as you can see that these are again spatially explicit, that is for every location we have a different set of coefficients. And apart from that there is the uncertainty which is through the these W variables. Now, this the weather component, it generates the variability and the and spatial correlation via the multivariate normal Gaussian process.

So, this is that like this W_N, W_X, this is considered to be following some kind of a Gaussian process which adds the randomness. And also, while adding the randomness, it also preserves the correlation between the different locations with the, like using its suitable covariance function whatever is used.

(Refer Slide Time: 24:54)

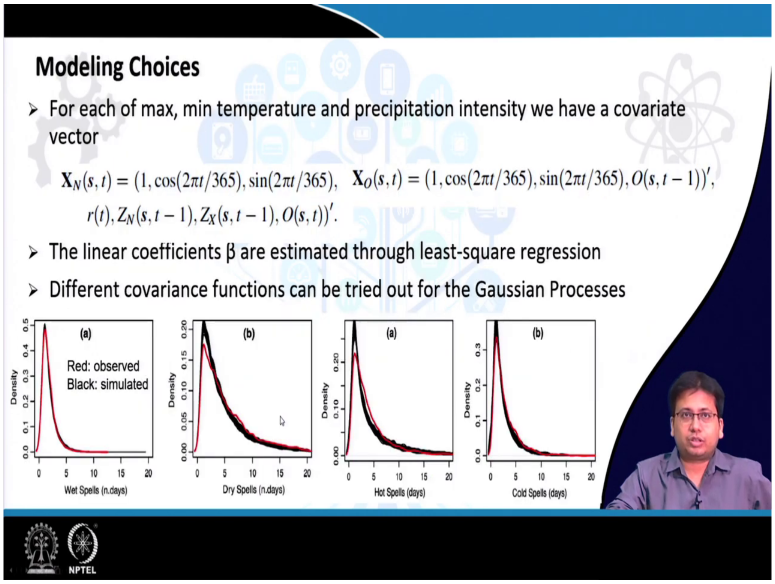


Now, coming to the part of the rainfall, the stochastic model for the rainfall. Now, this once again like the previous one; it is, first of all we decide whether the rainfall will occur or not at a given day. And if so, then what will be its intensity or amount that is the A. So, first of all the binary variable $O(s, t)$ that is like that follows a probability distribution and then where the latent process is a Gaussian process.

So, like there is a, so W_N, W_X these are of course, Gaussian process related to the temperature. Now, related to the occurrence of the rainfall there is another Gaussian process called W_O. Now, which on like that its values are sampled for every day at every location. Now, if that value crosses a threshold, then we interpret it to mean that it is going to be a, like I mean rainfall will happen at that location on that day.

And then, that is how we decide whether a given day is going to be wet or dry. And then, if it is a wet day then the amount of rainfall $A(s, t)$ is sampled according to some other distribution like this. So, and that also has a this covariance function which, like which helps it to maintain the correlation between the rainfall amounts across different locations.

(Refer Slide Time: 26:21)



Now, if we like come to the different modelling choices, there are like all these model parameters to consider all the covariates and so on and so forth. So, among the covariates, one thing which you may note is that it is really the year of the, I mean the day of the year which is considered to be an important covariate. I mean it is really the most important component of this maximum and minimum temperature is considered to be just the season itself.

That is on certain days we expect the temperature to be high, on certain other days in may be certain months the express expect it to be low and so on. So, all those things are presented as covariates. But not directly, but in like as you can see, like they are converted into some kind of a geometric function. This is just to like indicate the their periodic nature.

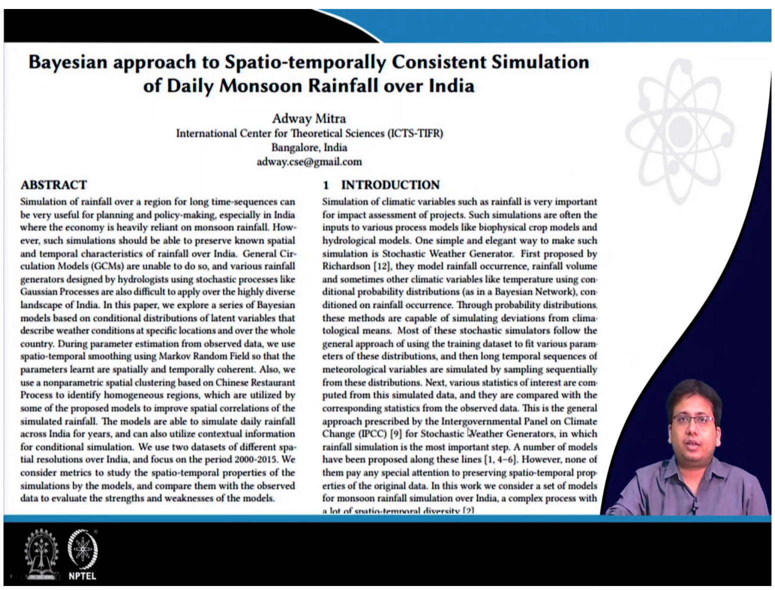
So, the temperature variation across the days of a year is considered to be, like it is represented as something like a sin curve and like, so that is why the day of the year is considered, like it is represented in this particular way using the geometric functions. And apart from these also, like the that whether it rained on the previous I mean what was that temperature on the previous days, what was the maximum minimum temperature, did rainfall occur or not.

These are all factors which like which are considered as covariance for the rainfall. Well, I mean, so sorry I mean for the maximum temperature. While for X_O, the which is this is the Gaussian process which on which it is decided whether rainfall will occur or not. So, for that also like we have the similar variables because the rainfall is also a seasonal thing. That also, like we can I mean the occurrence of rainfall can be treated as something like a sinusoidal variable. That is why we employ these geometric things.

And so, this is how the simulated statistics compare with the observed statistics. So, here they are considering the length of the wet spells, the lengths of the dry spells, lengths of the hot spells. This is related to the maximum temperature. Then, of the cold spells, this is related to the minimum temperature and so on. And you; so, the red is what is the observed distribution of the this spell lines and black is what is obtained from the simulations.

So, as you can see in all 4 cases, the red and the black almost coincide, which means that for this particular thing that for this particular quantity that is the length of spells, the simulation is near perfect. Similarly, other thing, other quantities can also be compared to the I mean to the observations, and they are also, they find a good match between the what is observed and what is simulated.

(Refer Slide Time: 29:17)



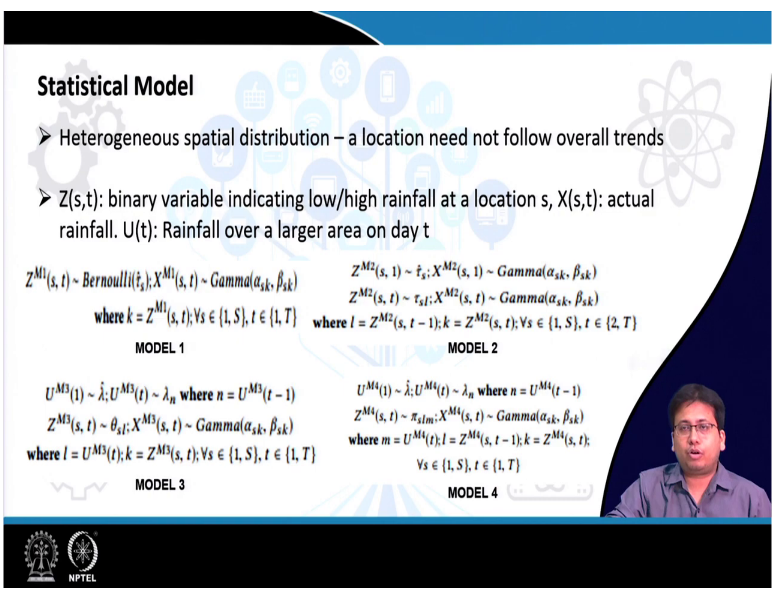
Now, we come to another paper, this is specifically for the Indian region. So, the last two papers we focused on a rather small homogeneous region. Now, let us consider a larger heterogeneous region like India, and we are trying to like simulate the spatio-temporally, like all the spatio-temporal properties of daily monsoon rainfall. So, the here with the simulation of rainfall over a region for making for long time sequences can be very useful for planning and policy-making, especially in India, where the economy is heavily reliant on monsoon rainfall.

However, each is have, such simulation should be able to preserve known spatial and temporal characteristics of rainfall over India. General circulation models are unable to do so, and various rainfall generators designed by hydrologists using stochastic processes like Gaussian Process are also difficult to apply over the highly diverse landscape of India. So, the heterogeneity is something which becomes a big challenge when we are considering a such a wide and diverse region.

In this paper, we explore a series of Bayesian models based on conditional distributions of latent variables that describe weather conditions at specific locations and over the whole country. Here during parameter estimation from observed data we use spatio-temporal smoothing using a Markov Random Field so that the parameters learned are spatially and temporally coherent.

So, this concept of getting spatial and temporally coherent values from Markov random field, we have discussed several times earlier. So, the same concept is used here also. And so, let us come to the models right away.

(Refer Slide Time: 31:11)



So, there are several models each in a particular way. So, here there; so again there is a two, it is a two-step process. There is a $Z(s, t)$ which is a binary rainfall indicating not whether rainfall will happen or in the earlier cases like in using this probit, the occurrence was considered as a binary variable which indicates whether rainfall will happen or not. In this case, it is not really that like that, it is more like whether it will be low rainfall or high rainfall.

And now what is low and what is high may vary from one region to another because we are considering a highly heterogeneous region. And $X(s, t)$ is the actual quantity of rainfall. And similarly we have another variable called U, which indicates the rainfall over the entire region. So on; like U is also a binary variable which like which means that; so X is or sorry Z is specific to one particular region, one particular location and U is specific to the entire region. So, in one model each of the locations are handled independent of all other locations. Each day is also handled independently.

In another model, we take into account the temporal coherence. That is it is more like a Markov process. The Z of a particular, the Z at a particular location and at a particular day depends on the Z on the previous day at the same location. And accordingly, we have the; once we have the Z accordingly we have the, we sample the values of the X also accordingly.

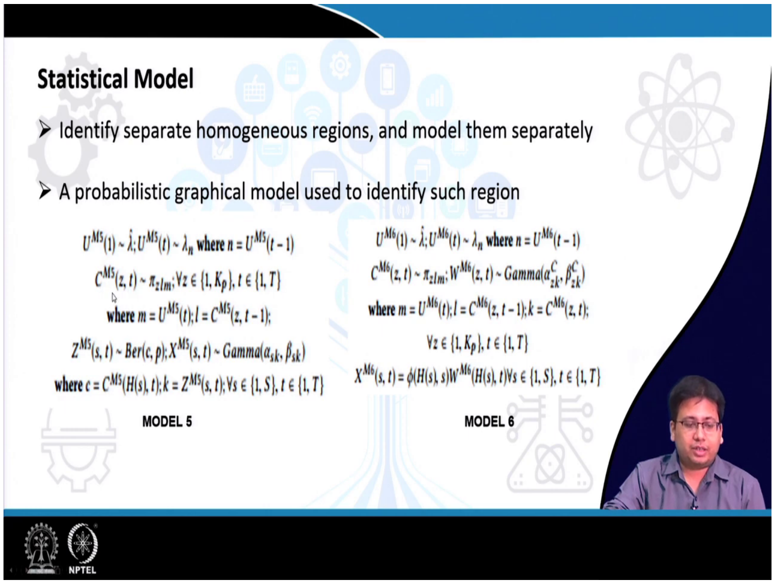
Like in another model, this Z it actually depends on U. That is on a given day whether it will rain, there will be low rainfall or high rainfall at a location, depends on its, depends on the rainfall conditions over the entire region. So, the relation between the entire region and a particular location that is somehow characterized by the probability distributions.

Now, there are like, when we are dealing with a large heterogeneous region like India, so there are certain regions like for example, the North East which often like acts in an anomalous way. Whenever the rest of India, in those times when rest of India is getting heavy rainfall usually the or often the North East does not get so much rainfall and vice versa.

On the other hand, there are certain regions in central India which get rainfall on those days when the like the region as a whole is getting a lot of rainfall. And similarly, on the days when the region on the whole is dry that region also remains dry. So, these kinds of local-to-global relations are captured through or this model tries to parameterize those relations.

And then all of these things can be done simultaneously, that is taking the temporal, spatial, and this local to regional effects at a time, like all of them can be handled.

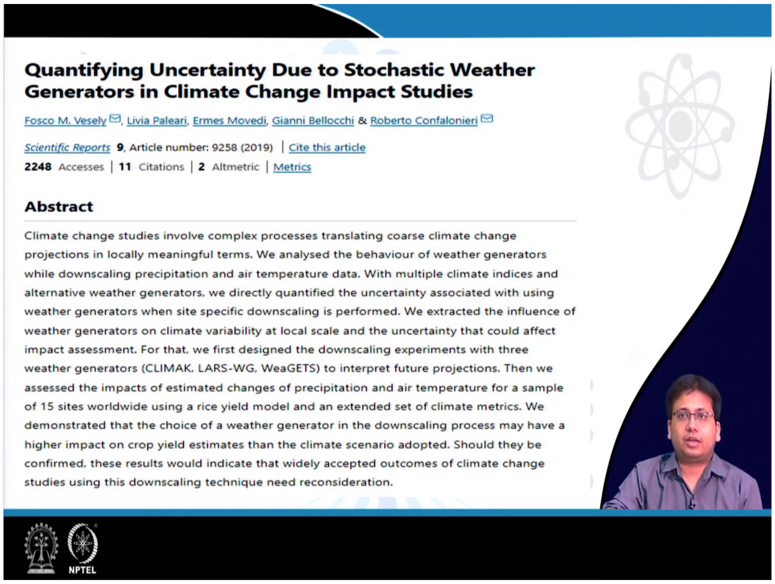
(Refer Slide Time: 34:14)



Different approach is to this we have already said that this is a large heterogeneous region, so why not try to first divide it into homogeneous regions and then like simulate the rainfall in all the homogeneous regions not independently, but with, but like there will be some overall governing variable that is the U basically. Which; the U which governs there which indicates the kind of rainfall over the entire region, that will have a, like have a bearing on each of the smaller homogeneous regions as well.

So, the first task is to divide the whole region into such homogeneous sub-regions and that is again achieved with the help of a Markov Random Field. So, this C is that region indicates the region variables. And the rainfall amount is then, at a particular location is then simulated based on the conditions in that homogeneous region C.

(Refer Slide Time: 35:20)



And so, now, there are certain; so, there is a huge volume of literature based on this stochastic weather generator and so on. And now, these are often used to quantify the uncertainty due in the climate change impacts, that is typically these are used these kind of generators are used in the future to run say in several years into the future and then try to study the nature of uncertainties and so on under different scenarios. That is as you have seen the these rainfall or these weather generators they take various covariates.

Now, some of these covariates may be coming from outside the model. So, that is the model itself takes care of only these 3 variables, the minimum temperature, maximum temperature and rainfall. But apart from that also, there can be other meteorological variables which are extraneous to the model.

So, like we in a sense, they can be considered to be forcing. So, one way to use these models is as follows. Say suppose we use these large scale simulation models like GCMs to simulate the climatic variables into the for many years into the future and so on.

And use those some of the variables thus simulated as extraneous variables to drive these kind of rainfall. I mean these weather generators and see the effects. Now, and this paper actually adds a note of caution to that. It says that the results, we can get as a result of this exercise may vary highly depending on the on which stochastic weather generator model which that we are using. One model may give one kind of result, another model may give a very different kinds of result.

So, if you are planning to use these for understanding future scenarios, it is necessary to exercise caution because these like basically because the different models of this type, they may not agree with each other, especially under the changing climate.

(Refer Slide Time: 37:36)

So, these are some of the different papers which we discussed today. That brings us to the end of this lecture.

Thank you. We will consider further applications of Machine Learning in Earth System Processes in the coming lectures. So, till then bye.