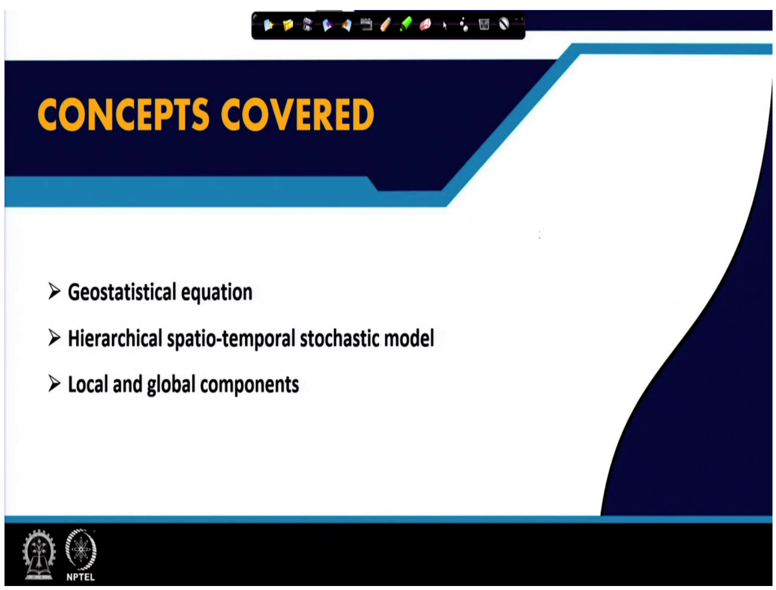
Machine Learning for Earth System Sciences
Prof. Adway Mitra
Department of Computer Science and Engineering
Center of Excellence in Artificial Intelligence
Indian Institute of Technology, Kharagpur

Module - 01
Spatio - Temporal Statistics
Lecture - 03
Geostatistical Equation for Spatio-Temporal Process

Hello. Welcome to the lecture 3 of this course on Machine Learning for Earth System Science.

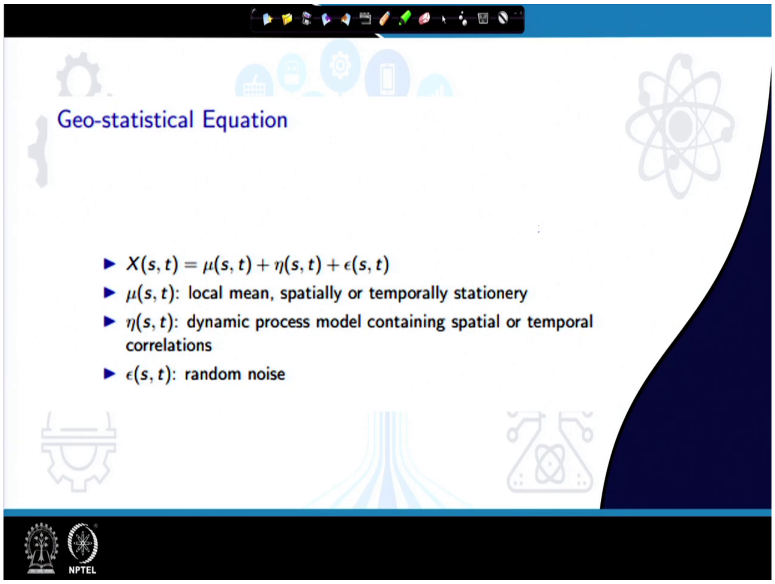
We are currently doing the first module which is on Spatio-Temporal Statistics. Today's lecture is on Geostatistical Equation for Spatio-Temporal Processes.

(Refer Slide Time: 00:43)



In this lecture, the concepts that we are going to cover are the geostatistical equation, the hierarchical spatio-temporal stochastic model, and local and global components.

(Refer Slide Time: 00:53)



So, let us start with the geostatistical equation. So, let us consider that we have some observable spatio-temporal variable X which may be any geophysical variable like say temperature, precipitation, wind speed or anything you like which is may be suitable.

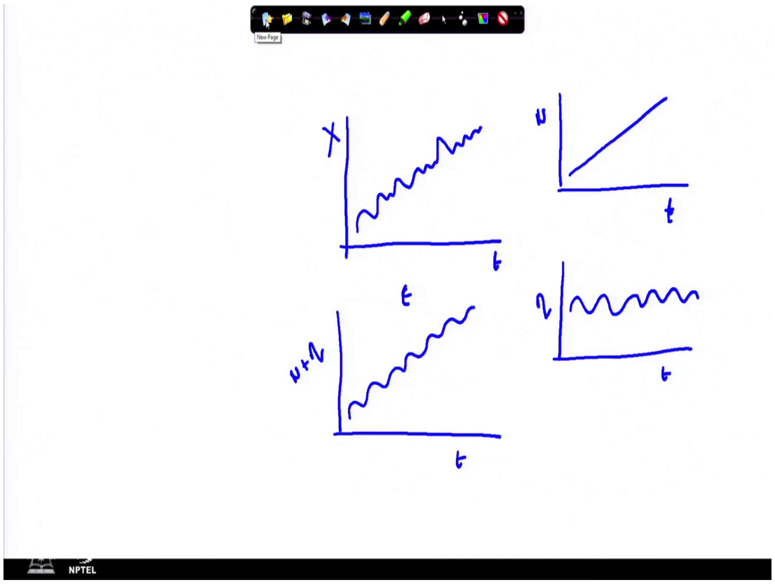
Note that this is a spatio-temporal variable, so it has a spatial index s and a temporal index t. Now, this the value of this variable $X(s,t)$ at any given location s and time point t, like for the purpose of understanding we will decompose it into 3 components.

The first one is the local component or μ, the second one is the dynamic or the like you may call it as the global component η, and the third one is the random component which is ε. So, each of these, the first one the μ or the local component this is often a spatially or temporally stationary process, and it like, it may be considered as the function of that particular location or that particular time.

Now η(s,t), like this is something which like this is a dynamic thing which may vary over space or over time. And the last one which is ε this is considered to be the IID noise, that is it is independent and identically distributed at every location and every point of time.

So, let us illustrate this matter with an example.

(Refer Slide Time: 02:27)



So, let us consider a time series of a random variable X. For the time being we are considering only one particular location, ok.

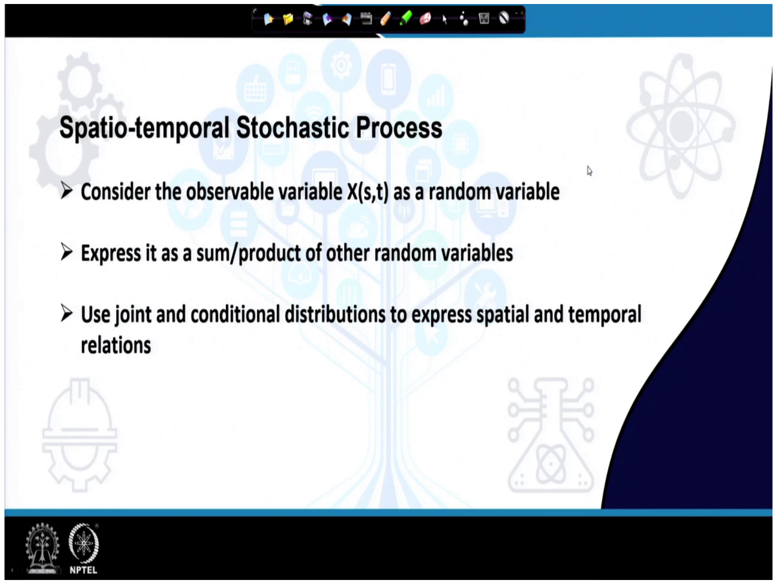
So, let us say that the observation of $X(t)$ over time looks somewhat like this, ok. And so, now, like, so those of you who are familiar with time series analysis will note that this kind of a time series can be broken into 3 components.

The first one, we can call as the trend, which is like you can say that the it is generally increasing or monotonously increasing over time, right. Then, the secondly, there is a periodic component also, periodic or cyclical component which we may call as η, which just shows this kind of a cyclic or periodic behavior.

Now, if you superimpose this μ, the trend and the periodic behavior η like this, then the signal we which we get is somewhat like this. But note that this is different from the original signal where like the difference is that sometimes may the like the while this one has uniform peaks and troughs. In this case, sometimes the peaks are high, some troughs are low and things like that. So, this discrepancy is taken care of by the by that noise which is the ε, right.

So, this is just an example. So, this concept can be further moved to the domain of the spatio-temporal process also, ok.

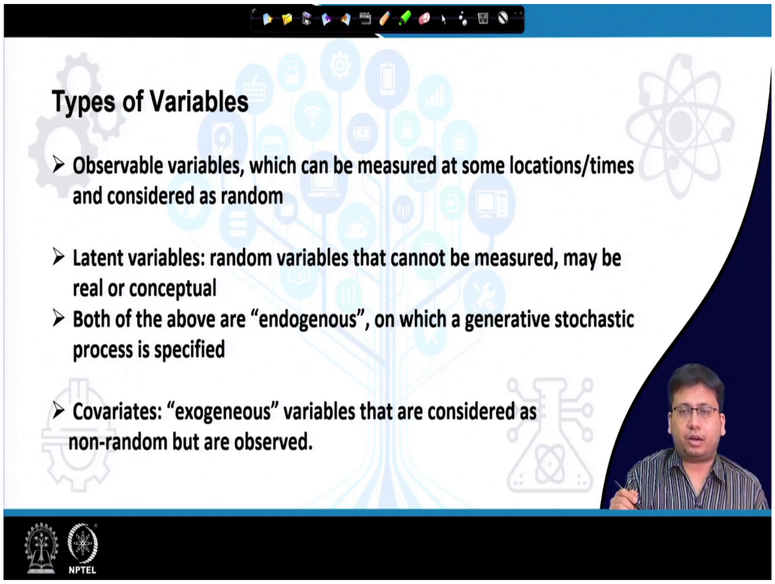
(Refer Slide Time: 04:10)



So, now let us consider the Spatio-Temporal Stochastic Process. So, where the consider the observable variable $X(s,t)$, which we considered earlier, this variable $X(s,t)$, let us consider it as a random variable.

Now, we will express that random variable as the sum or product of other random variables. And we will use joint and conditional distributions to express the spatial and temporal relations between the different random variables.

(Refer Slide Time: 04:43)



So, in the stochastic model which we are going to consider, based on the geostatistical equation will contain different kinds of variables. First of all there are the observables, so these are the variables which can be measured at some location or time point. And they are considered as random variables. The observations which we have of these can be considered as instances or of the random variable.

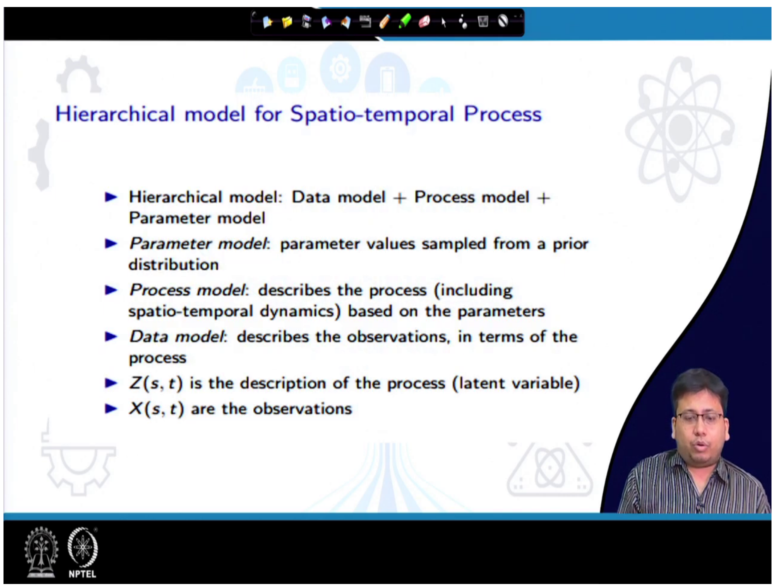
The second type of variable are the latent variables. So, these are also random variables, but they cannot be measured. So, and these latent variables, they can either be real variables that is it may be simply some kind of a physical variable which is difficult to measure due to lack of suitable instruments or and so on. Or they may be purely conceptual variables, like for example, we may define something like the state of the atmosphere kind of thing.

Now, the state of the atmosphere that is obviously, some like a conceptual quantity there it, it does not refer to anything, any like any physical quantity in particular, but it might be used as some kind of a random variable in the models that we that we will consider.

So, both the above variables, the observables as well as the latent variables are endogenous variables or which means that the model will specify some kind of a generative stochastic process on them. That is we will specify from what kind of probability distributions these variables will be are following.

Apart from that there are also some covariates which are exogenous variables which means that they are non-random variables, but we know their values. That is they like you can consider, by the word exogenous basically means something that has come from outside. These are not variables whose origin is explained by the model. We simply consider that they are provided as side inputs to the model.

(Refer Slide Time: 06:47)



Now, the when we come to the model of a spatio-temporal process, I mean the stochastic model for a spatio-temporal process that such a model is hierarchical in nature. So, what is the hierarchy? So, it is a combination of a data model, a process model and a parameter model.

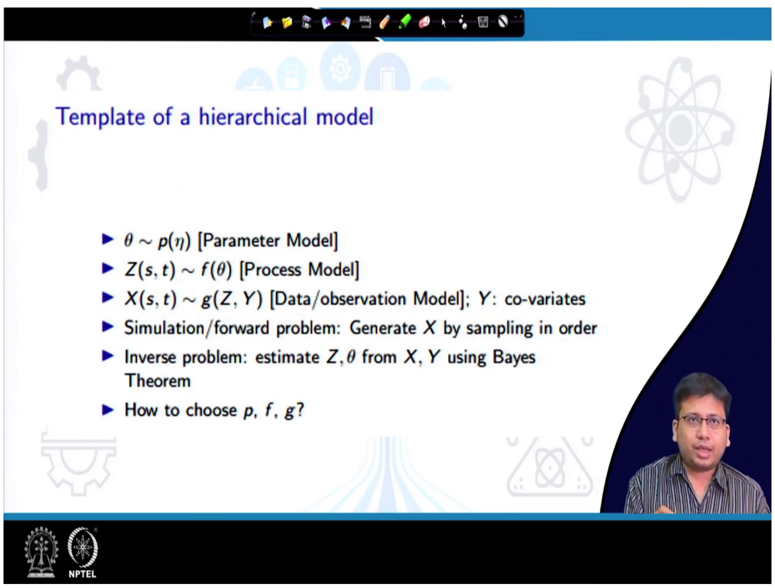
So, parameter model, first of all like that lies at the lowest level of the hierarchy. So, these are basically parameter values which are sampled from some kind of prior distribution.

Next comes the process models. So, the process model somehow describes the process that we are trying to model using its spatio-temporal dynamics using the parameters. So, this process model again, will like it is it usually explains the generative parts of the latent variables like using the parameters as the para mentioned these; so, the parameters that were obtained from the parameter model those parameters are used as the parameters of whatever probability distributions these those latent variables will be following.

And finally, comes the data model. So, the data model basically describes the observations in terms of the process. So, $Z(s,t)$ is the description of the process and like these refer to the latent variables and $X(s,t)$ are the observations. So, note that Z and X in this case need not be individual variables, but it may also be collections of variables. That is like $X(s,t)$ need not refer to only a single variable, but it may refer to a collection of variables like say temperature precipitation and wind speed at a given location and time.

Now, whether it refers to one variable or to a collection of variables will become clear from the context.

(Refer Slide Time: 08:44)



Now, here is a template of such a Hierarchical Process Model. So, as I mentioned, first of all there is the parameter model. So, let us say that θ denotes the parameters.

Again, it may not be a single parameter value, it may be a collection of parameters. And let us say that they follow some kind of a probability distribution p and which might be something like a Gaussian distribution, Beta distribution, Gamma distribution, Bernoulli distribution depending on the nature of these parameters θ.

And those probability distributions will again have their own parameters, like we know the Gaussian distribution has the mean and sigma parameters, the Bernoulli distribution has the p parameter and so on. So, like those kinds of parameters of these prior distributions those are known as the hyper parameters. So, they are denoted by this η.

Next, once we get the parameters by sampling them from this kind of a probability distribution p, those parameters θ are used as parameters of the process model. So, the process model basically gives us the latent variables Z. And these the latent variable they follow some kind of a probability distribution f using the θ, as their as parameters. So, once again this f can again be a Gaussian distribution or a Gamma distribution or whatever it is, depending on the nature of the problem. And accordingly it needs the parameters θ.

Now, once we have got the latent process variable Z, next comes the data model where we actually try to generate the observations X, the $X(s, t)$ we put some kind of a stochastic model on the observations and these of the they this kind of a stochastic model takes as its input the latent variable Z which have already been generated by the process model and additionally it also takes the covariates Y.

So, these covariates as we just mentioned are the exogenous models. So, note that unlike the $Z's$ which are the endogenous models, the endogenous; we are calling them endogenous because we have actually put some kind of a generative model on them here, in the process model. But in case of Y, there is no such thing. It; like you can say it makes a lateral entry into the data model at this stage, ok.

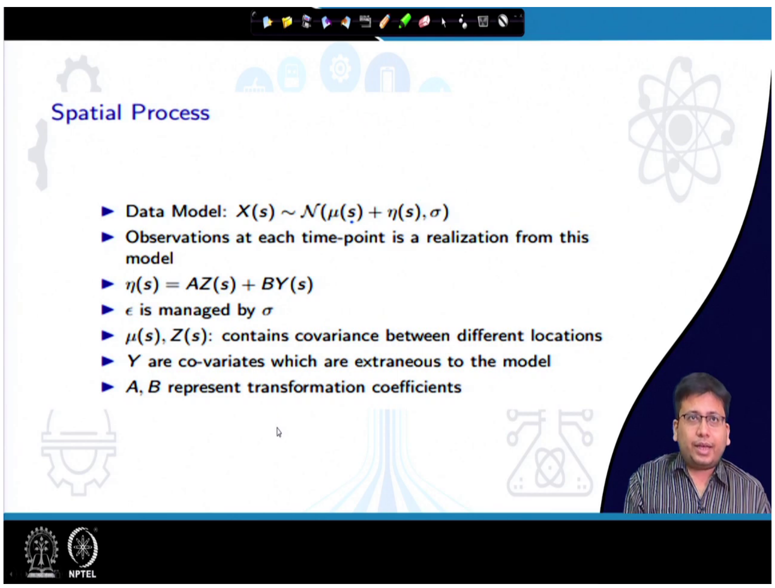
So, now this kind of a model it works in two modes, one is the simulation mode or the forward problem where you generate the X by sampling in order, that is first you sample the values of θ from the distributions p, then you sample the latent variable Z using θ and then you finally, sample the observe observable variables X using Z and Y.

This is the like in, like you can say that in this mode you are actually creating new observations. the like the So, if you are interested in doing some kind of simulation of the future or what future values of that variable may look like, then you may use this forward model.

The more standard thing to do is the inverse problem where you already have the observations X and maybe also you have the observations of the covariates Y. Your aim is to estimate Z and θ using the Bayes theorem.

And then of course, there are like there are design choices like how to choose these distributions p, f, g and how to like specify these hyper parameters η. So, these are the last set of questions are to be answered by the model designer. And the choices they will make will determine how good the simulation model is going to work.

(Refer Slide Time: 12:47)



Now, this can be used in the spatial domain or in the temporal domain or both simultaneously. So, let us say that we consider a spatial process. So, for now let us forget the temporal component all together. Let us just drop the t and focus only on s.

So, here the idea is that at every location s, we have the particular observable X and we decide whoever is designing the model decides that X follows a Gaussian distribution. So, Gaussian distribution, it will like it will require the mean and the variance. So, let us now, for now just assume that this Xthis is just a real number. So, its mean will be a real number, its variance will also be a real number.

So, this variance σ which we are putting is basically just the manifestation of the or sorry we can say is it is just a parameterization of the random noise which we talked about the ε of the geostatistical equation, ok. And the $\mu(s) + \eta(s)$ that is the mean of this Gaussian distribution that is the non-random part of the geostatistical equation, ok. So, the like in this version the observations at every time point is a realization from this model.

So, that is this model is specific to spatial location, but like, it is IID with respect to time that at any given location the observations follow the same distribution, but at different location they may follow different distributions. Because μ and η they are also spatially like index by space, that is they vary from one location to another.

Now, what about these components μ, η etcetera? So, μ is the local component, so we can assume that μ is fixed for a particular location. Now, η on the other hand like we can call it as something like as the dynamic or we can call it as the global component. So, η is actually like you can say that, it is in a sense it is dependent on the latent variable at different locations and or it may also be dependent on the covariates at the different locations.

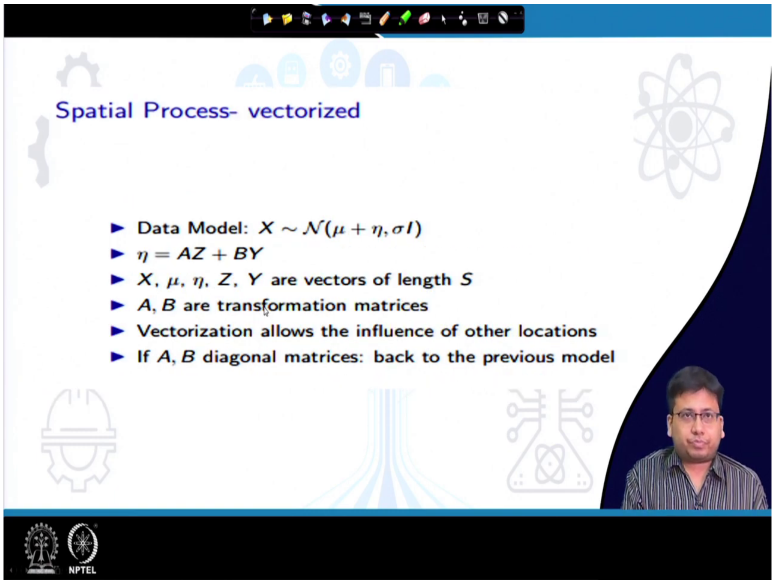
And this the this Z, it must be developed in such a way or it may follow such a distribution that it should be able to encode the covariance between the different locations.

So, we know that like we had already studied the concept of auto correlation. So, we know that this observation X at different locations they may be autocorrelated.

In fact, we considered the auto regressive model in which we tried to express X at a location $s1$ in terms of X at other locations say $X_{s2,s3}$ and so on with the help of a regression model in lecture 2. So, that basically means that there is some kind of covariance between these X at the different locations.

Now, that kind of covariance will not come if we like model s separately or independently at different locations. For that purpose some kind of spatial dependency has to be brought in, some kind of covariance has to be brought in somehow. So, that is usually managed through this component Z. And Y as we know are like these are the the extra the exogenous covariates. And these A and B, these you can say these are the transformations coefficients.

(Refer Slide Time: 16:51)

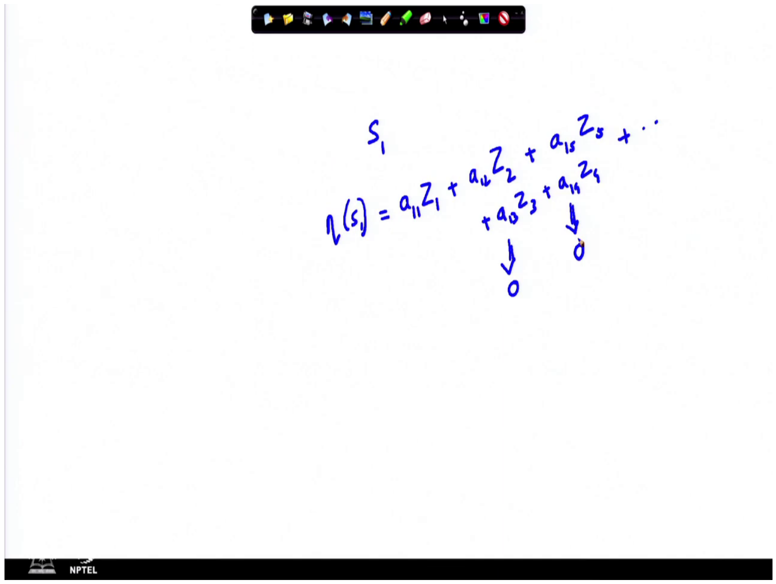


Now, this same thing we can consider it in a vectorized sense also. So, instead of considering a location by location like this, like we can consider the data model as or the variable X as a vector, that is a vector over all the locations. And in that case we just get a vectorized notation of the same thing. So, instead of considering the like an univariate Gaussian distribution, we now consider a multivariate Gaussian where the like the mean component is a vector and the instead of a variance we actually have got a covariance matrix.

Now, in it might make sense to use the covariance matrix as something like as isotropic that is as σI because the σ is basically nothing but the observation noise, the random noise, the ε of the geostatistical equation which may be considered to be identical at all locations. So, that is why we can we do not need to make it as location specific. We can just use the identity matrix in this case.

Now, the mean vector μ and η, $\mu + \eta$ these two things they should be like location dependent. Now, this η once again we will have to express it as $AZ + BY$, where Z is the latent variable at the different locations and Y are the co variates at the different locations. And η is like again is a vector which combines these the latent variables at different locations and the covariates at different locations with the help of a this kind of a of transformation matrices.

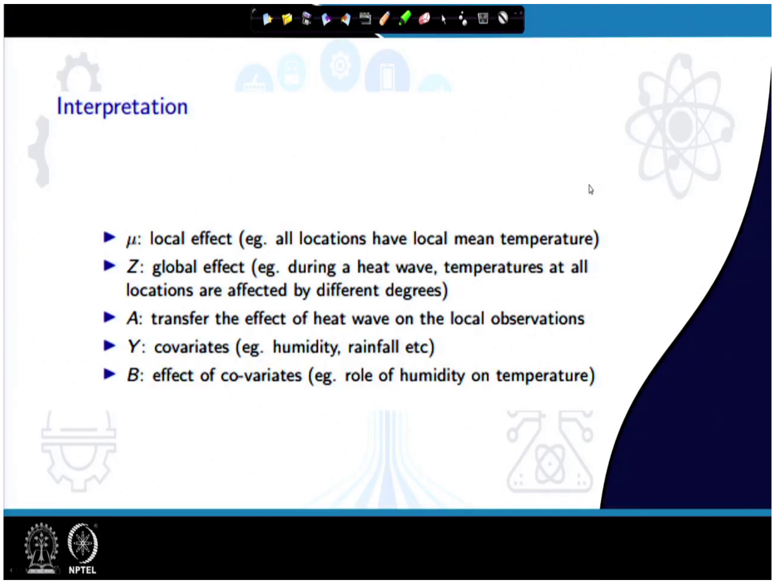
(Refer Slide Time: 18:46)



So, we can say that at any given location let us say $S1$, so the $\eta(s1)$ may be expressed in terms of the $Z's$ of other locations.

Let us say $a_{11}Z_1 + a_{12}Z_2 + a_{15}Z_5$ and so on. I mean, I have dropped 3, 4 and others, we can just to a just to allow for the possibility that the at any particular location $s1$ it is possible that the not all the locations influence. I could have also written it in this way $a_{13}Z_3 + a_{14}Z_4 + a_{15}Z_5$ etcetera, where some of these coefficients a they may go to 0, ok.

(Refer Slide Time: 19:44)

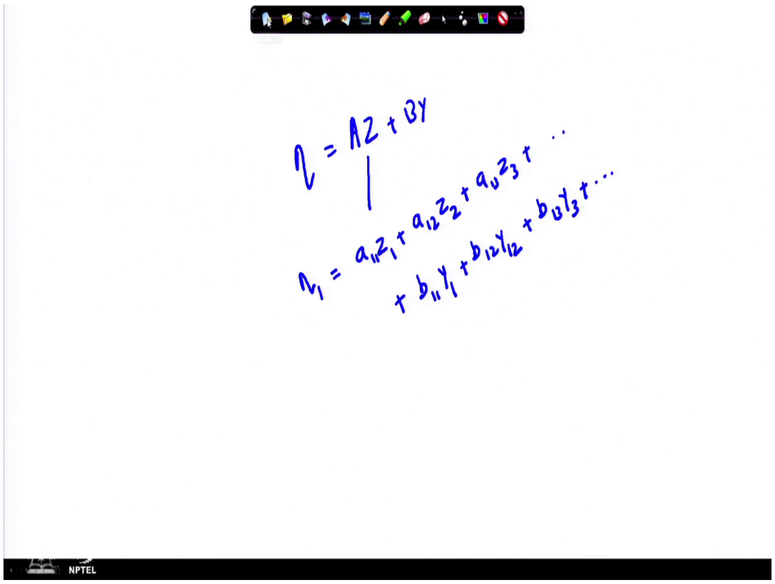


So, the interpretation of this is as follows. μ is the like the local effect where all the exam for example, all the locations may have some local mean temperature. Z is you can consider as some kind of a global effect that is during the heat wave the temperatures at all the locations are affected by different degrees. And a is somehow the transferring the effect of the heat wave on all local observations.

So, so just, so like when I expressed $Z1$ in terms of the Z at other locations, I mean the Z at the first location is impacted by the $Z's$ at locations $s2, s3$ etcetera. So, the what this a basically does is transfers the effect of Z at other locations to the Z at or rather to η at one particular location.

And then of course, we have the $Y's$, and similarly for the covariates also we have like we may want to like incorporate their effect of the covariates of at different locations into the η at a particular locations.

(Refer Slide Time: 21:04)



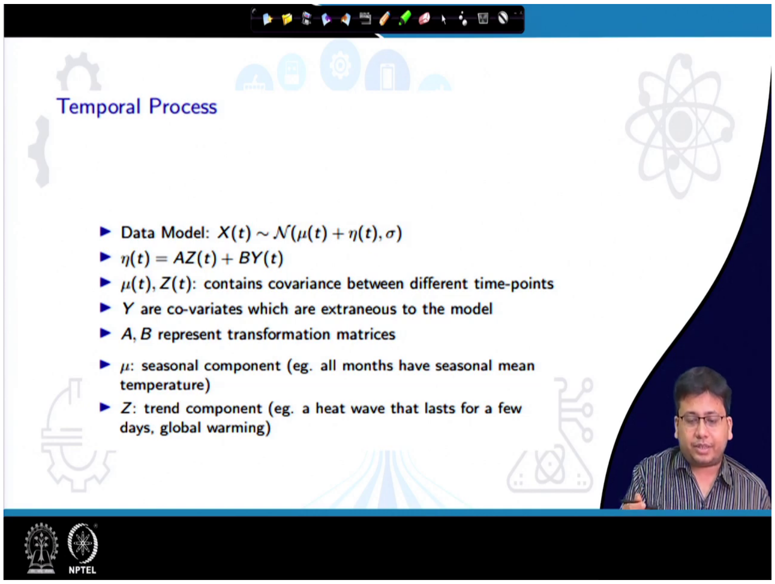
So, just like in the previous case we considered like this. So, where, so like $\eta(s1)$ or $\eta(1)$, if you want to call it this can be written as

$$\eta(s1)/\eta(1) = a_{11}Z_1 + a_{12}Z_2 + a_{13}Z_3 + \dots + b_{11}Y_1 + b_{12}Y_2 + b_{13}Y_3 + \dots \text{, right.}$$

So, like as you can see what these a and b, these coefficients are doing is that they are transferring the effects of Z and Y at different locations to the η of one particular location, ok. So, this vectorization, basically what it does is it allows the influence of the other locations instead of making the model location specific as it might have appeared in the scalar version.

Now, if this a and b are basically diagonal matrices that is let us say a only a_{11}, a_{22}, a_{33}, these are only non-zero and a_{12}, a_{13}, etcetera these are all zeros, then it is back to the original back to the previous situation where basically the η of every location is impacted by the Z and Y of the same location, ok.

(Refer Slide Time: 22:44)



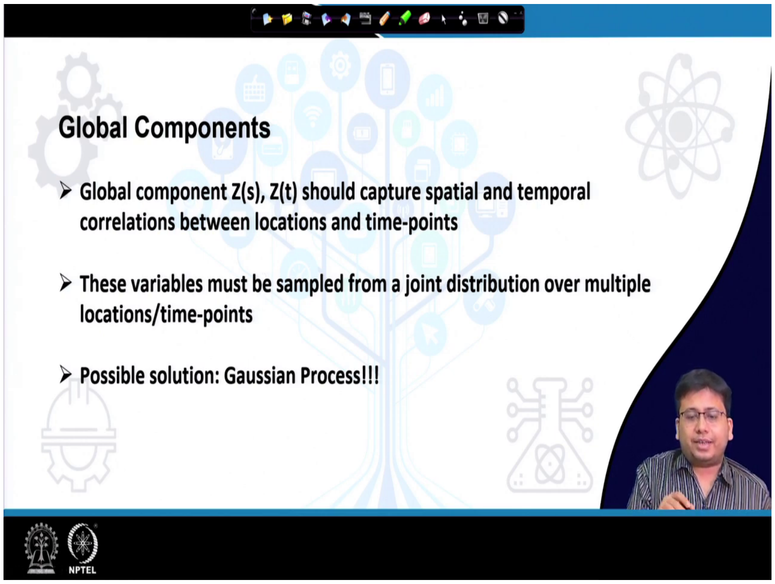
Like this, so we consider the spatial process with the same thing has a temporal analogy also, where instead of considering different locations we can also consider different time points. So, in that case, $\mu(t)$ and $Z(t)$ these should contain the covariance between the different time points rather than the different locations. And A and B continue to be the transformation matrices as usual.

So, in this case, the μ might be the seasonal, like one of them might be the seasonal component as I drew earlier. And the other might be the trend component which like which is the which may be expressed through the η, ok.

And the η, in this case is the encoding of some kind of a dynamical process which impacts like a longer particular time period. It might be something like a heat wave for that lasts for a few days and so on.

So, the Z is basically its like some kind of a dynamical process which a Z, basically Z is the latent variable, but the latent variable might be used to express some kind of a dynamical thing or a dynamical process and that impacts the observation X through this η. So, remember that this η is basically a function of this latent variable Z along with the different covariates.

(Refer Slide Time: 24:17)



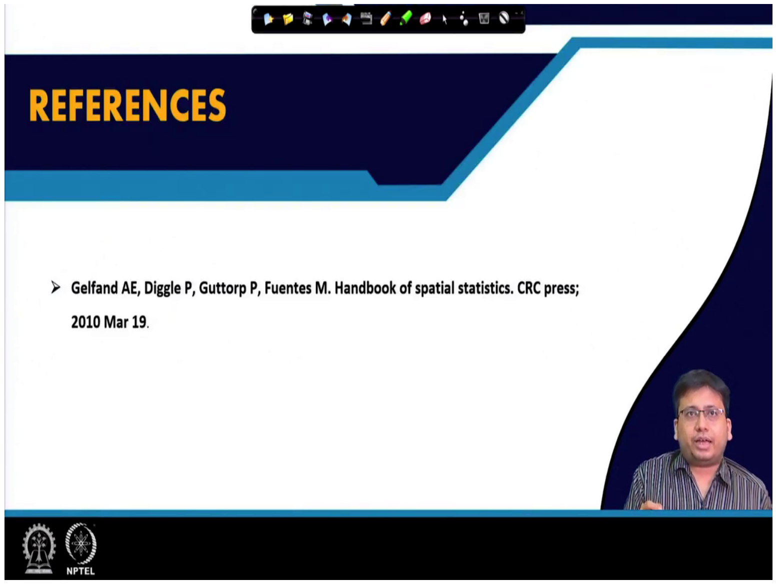
So, like the global components $Z(s)$ and $Z(t)$, they should capture the spatial and temporal correlations between the different locations and time points. So, in this case then we have the $Z(t)$, in the previous case we had $Z(s)$. So, we have to; so, whatever kind. So, the note that these are basically the latent variables for which we are defining some kind of a generative model in the second layer of the hierarchy.

So, they whatever stochastic model we are putting on $Z(s)$ and $Z(t)$, they should be in such a way as, so as to incorporate or express the spatial and temporal correlations. So, these variables they cannot be sampled I in an IID way, that is we cannot sample Z at different locations ideally like independently of each other. If we do that, then we will be will not be able to capture this kind of a correlation.

So, we have to sample them from some kind of a joint distribution over multiple locations or multiple time points. And one possible way in which the such a thing can be done is the Gaussian process.

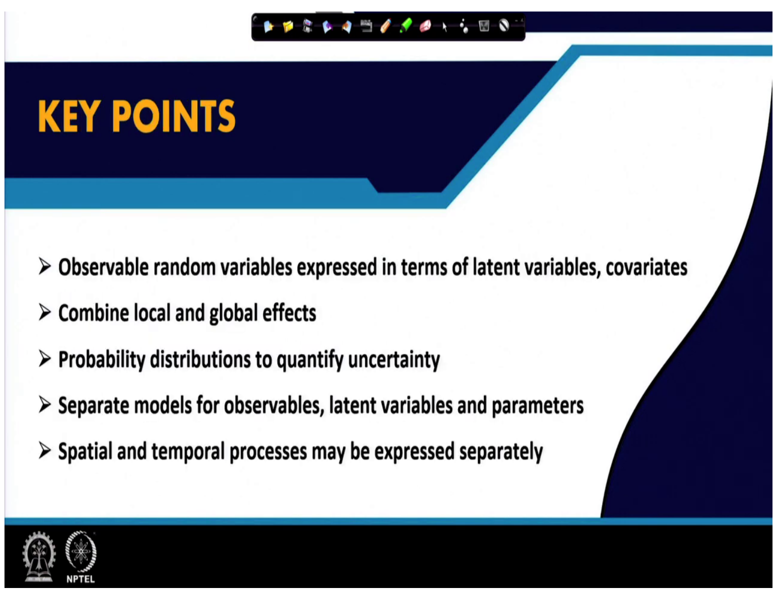
So, in the coming lecture, we will be focusing on this kind of Gaussian Process only.

(Refer Slide Time: 25:40)



So, that brings us to the end of this lecture. We will continue again in the next lecture which is on Gaussian Process.

(Refer Slide Time: 25:48)



So, the key points to take away from this lecture are as follows. First of all the observable random variables might be expressed in terms of the latent variables and covariates. Combine the local and global effects through the constructs like μ and η.

The probability distributions are used to quantify the uncertainty. Uncertainty at the level of the observations that is done through the random noise or other uncertainty in terms of the various latent variables which might be captured through the other in the process model.

And finally, we have separate models for observables, latent variables as well as the parameters. And the spatial and temporal processes may be expressed separately. Like like in this case as you see we had separate temporal process which focuses only on the time component. Before that we considered a spatial process which components only on the space component.

So, we might build a, when we are building a spatio-temporal model, we may build a spatial model and our temporal models separately and somehow combine them together. So, how that is done, we will see in the following lecture.

Till then good bye.