Course Name: Watershed Hydrology

Professor Name: Prof. Rajendra Singh

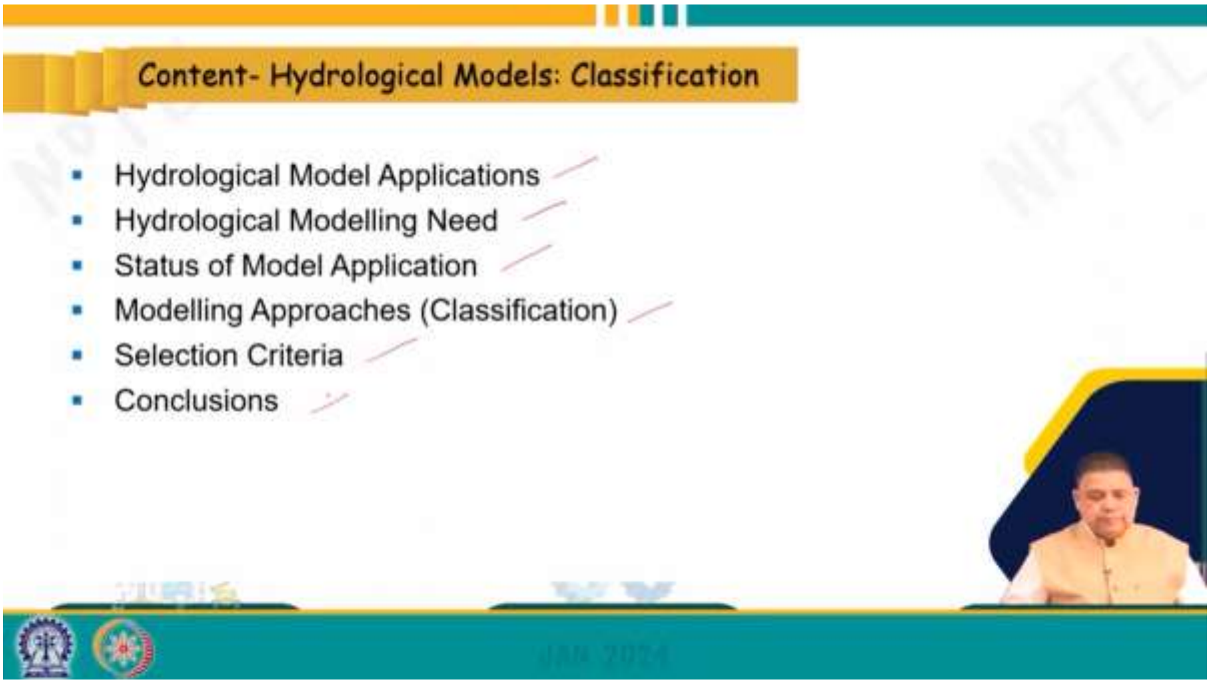
Department Name: Agricultural and Food Engineering

Institute Name: Indian Institute of Technology Kharagpur

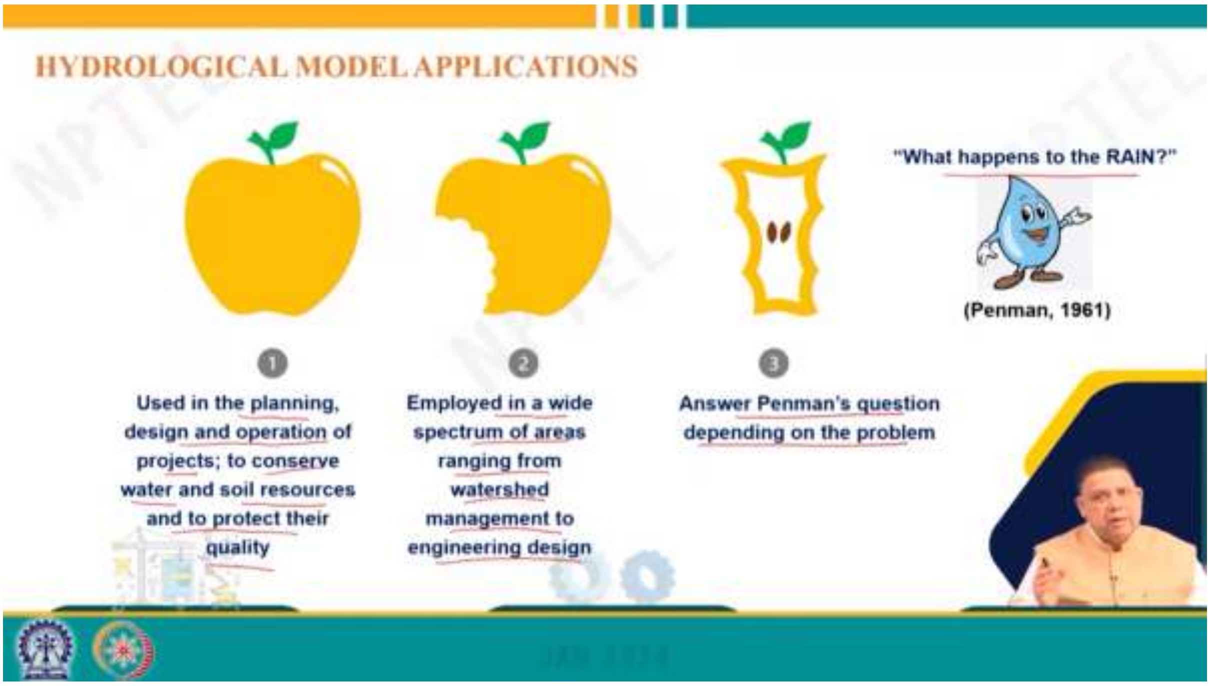
Week: 08

Lecture 37: Hydrological Models: Classification

Hello friends, welcome back to this online certification course on Watershed Hydrology. I am Rajendra Singh, a professor in the Department of Agriculture and Food Engineering at the Indian Institute of Technology Kharagpur. We are in Module 8, and this is Lecture Number 2, the topic of which is Hydrological Model Classification.



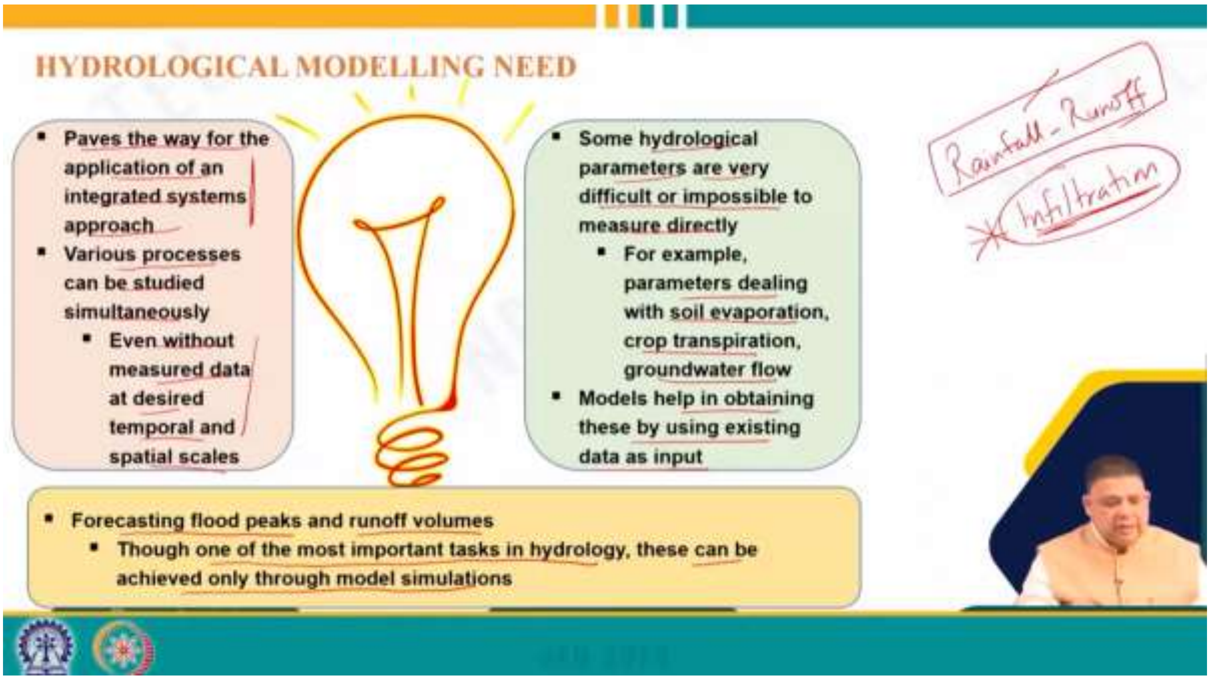
In this lecture, we will first talk about the hydrological model applications, the need in hydrological modelling, the status of model applications in various fields of hydrology, and then we will discuss the modelling approaches, classification model selection criteria, and finally, the conclusions.



Now, coming to hydrological modelling applications, these models are used in various facets of hydrological water resources planning and management. For example, they are used in the planning, design, and operation of projects to manage water and soil resources and to protect their quality.

So, as you can see, whether it is planning design or the operation of projects, it could be soil conservation, water conservation, or the quality aspects; both the quantity and quality hydrological models have been applied. They have been employed in a wide spectrum of areas ranging from watershed management to engineering design. So, obviously, whether it is just the policy matters or operational engineering design or project management, everywhere hydrological models have been applied in the field of water sources or in hydrology. And, of course, one of the major applications has been to answer Penman's question, depending on the problem. We remember Penman's question: what happens to rain?

So, basically, the rainfall-runoff transformation process has been one major area of application of hydrological models as far as hydrologic science is concerned. Now, why models are needed in hydrology is the next question.



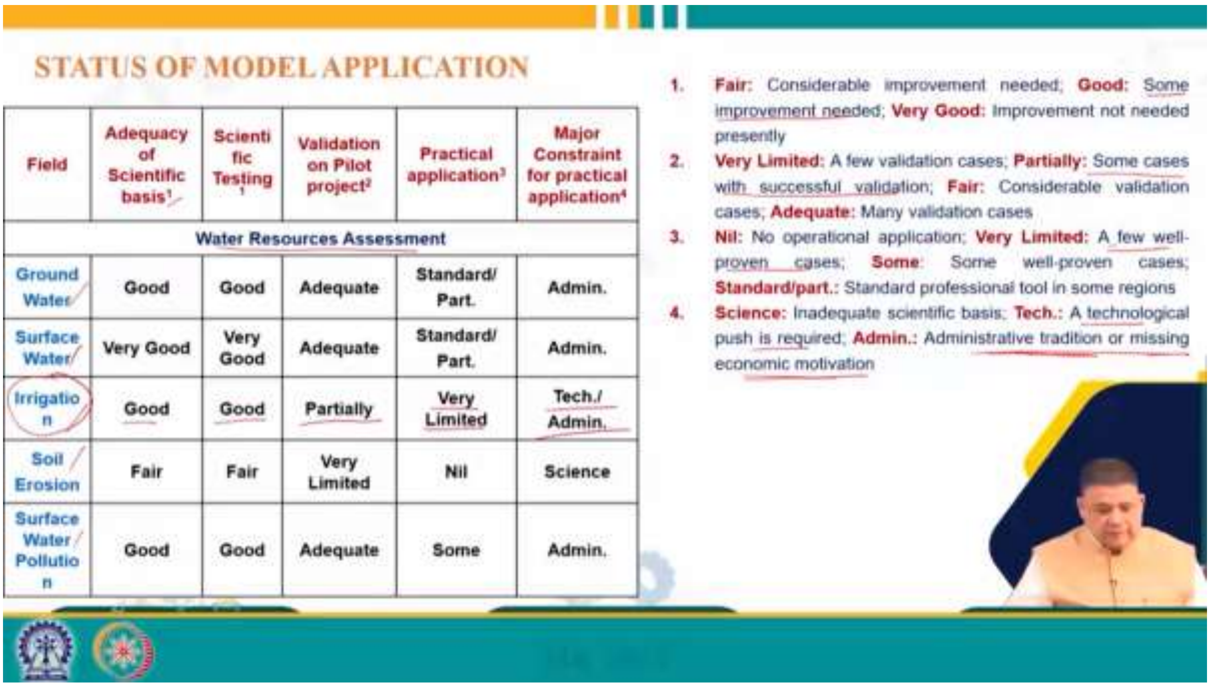
Hydrological models are needed in hydrology because they pave the way for the application of an integrated system approach, where various processes can be studied simultaneously even without major data or desired temporal and spatial skills. So, obviously, as we have already discussed, while formulating model formulation or while conceptualizing the model, the modeler has the freedom to use or neglect a particular process, for example, interception, say, for example. So, obviously, and also, on the other hand, you may not have, for example, you are trying to analyse rainfall-runoff transformation process, but you do not have soil characteristics or infiltration data, suppose you do not have.

So, obviously, you cannot put the infiltration process into your model because you will not have the soil characteristics or parameters pertaining to infiltration. So, that is why your model will not work. That is why you may not include infiltration in your model, but because your model is about the rainfall-runoff transformation process and we know that physically infiltration is a component. So, obviously, even without having the data on infiltration, you may have some idea about what is happening because you know there is rainfall, there is some interception, you have the data on evapotranspiration, suppose you have the requisite data, and of course, you are measuring the runoff at the end. So, obviously, because everything else is known, the remainder will be infiltration. So, basically, the water balance works. In any hydrological model, the water balance equation will work. So, obviously, if you do not have data on a particular aspect, you will be able to get some information about that aspect because the model helps us to carry out integrated systems modelling. It allows us to adopt the integrated systems approach, and that is why, as I said, even without measured data or desired temporal scale, you can obviously model that particular process. Also, at the same time, some hydrological parameters are very difficult or impossible to measure directly.

For example, parameters dealing with soil evaporation, though evapotranspiration includes a part of that as soil evaporation, but it is very difficult to really measure. Similarly, crop transpiration, groundwater flow, all these parameters, though they play a very significant role in hydrological modelling or the rainfall-runoff transformation process, are very difficult to measure. If you have some rough idea, then you can get an idea about the parameter values. So, models help in obtaining these by using and adjusting data as input. So, whatever data we have, we input because we will be modelling this process.

So, one or two parameters whose values were not known or measurable, you can try to get an idea about their values even though measuring them is not possible. Also, there are exclusive ways, like for example, forecasting flood peaks and runoff volume. Suppose you have to forecast a flood. So, obviously, you cannot carry out experiments on creating floods. In that case, the only solution is through modelling. Thus, one of the most important tasks in hydrology, forecasting flood peaks or runoff volumes, can be achieved only through model simulation.

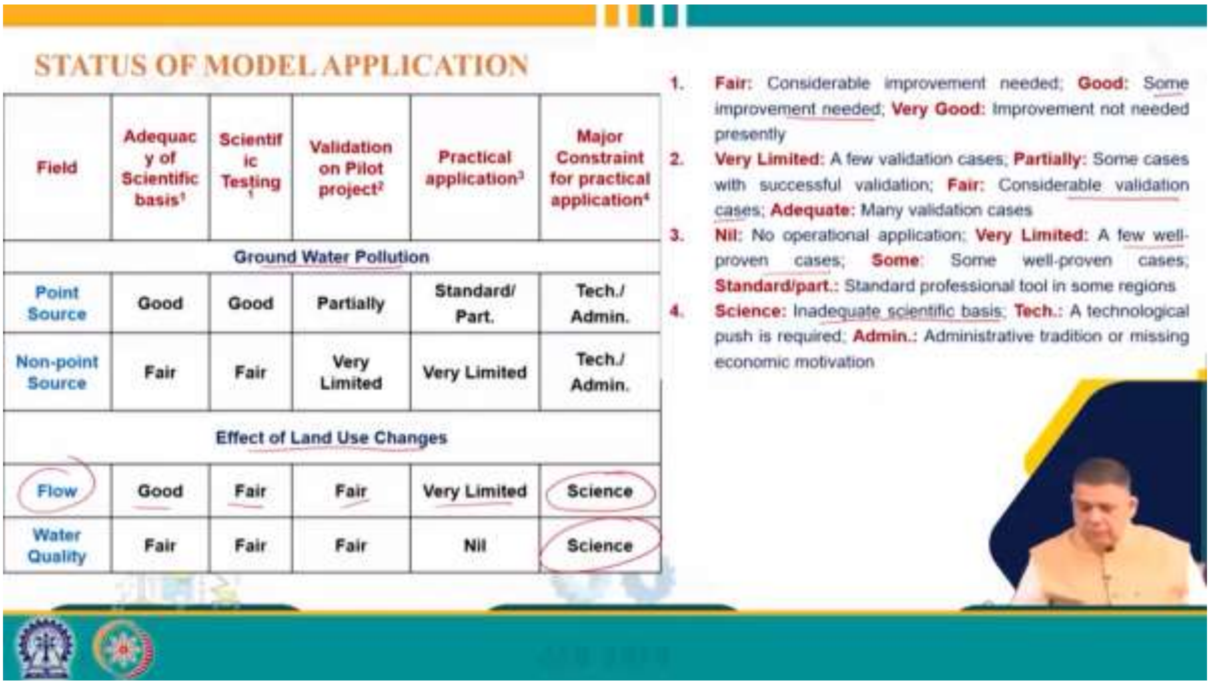
There are some exclusive fields where modelling is needed because we may not have data on certain aspects while modelling. In those cases, models can help. Similarly, for some parameters that are not measurable or difficult to measure, we can have a fair idea of those parameter values if we use models.



Now, we come to the status of modelling applications. This table shows the status of modelling applications in various fields. Let's take water resources assessment, groundwater assessment, surface water irrigation, soil erosion, and surface water pollution as examples.

For irrigation, for example, the adequacy of the scientific basis is good. When we say the adequacy of the scientific basis is good, it means some improvement is needed. It is not yet perfect; there are some components of the model that need improvement, but otherwise, the models are good. Scientific testing is also good, meaning that some improvement is needed, and we still need to test such models rigorously. Validation on pilot projects is partial. When we say partial, it means that some cases with successful validation are available around the world. This is the world status we are talking about.

Then, if it comes to practical application, it is very limited; that simply means that only a few well-proven cases are there, only a few limited cases are there. And when it comes to major constraints for practical applications, it's the technology and administration. Technology and administration mean a technological push is required, and administration means administrative tradition or missing economic motivates. So, obviously, as you know, in large irrigation systems, you have many gates. So, to operate them automatically, you need sensors. And so, obviously, the technology has to be developed yet. And of course, the administrative tradition and the missing economic motivation with the department officials; that is a major reason why the irrigation models have not become popular as far as field applications are concerned.



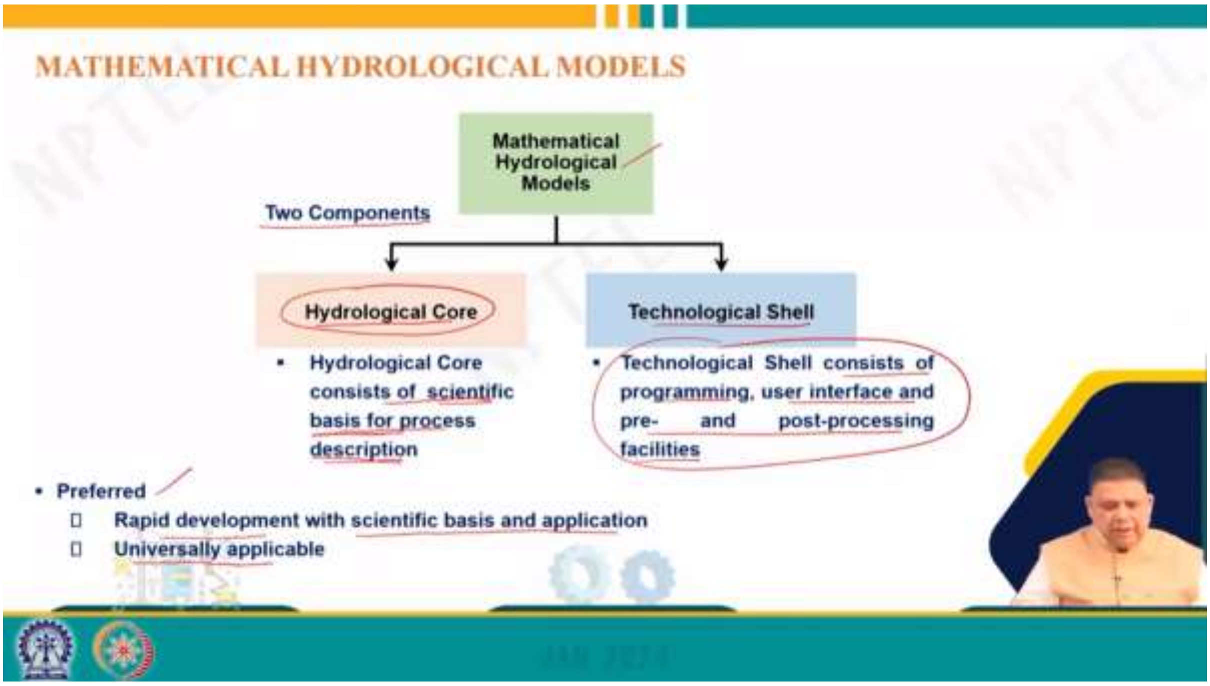
STATUS OF MODEL APPLICATION

Field	Adequacy of Scientific basis[1]	Scientific Testing[1]	Validation on Pilot project[2]	Practical application[3]	Major Constraint for practical application[4]
Ground Water Pollution					
Point Source	Good	Good	Partially	Standard/ Part.	Tech./ Admin.
Non-point Source	Fair	Fair	Very Limited	Very Limited	Tech./ Admin.
Effect of Land Use Changes					
Flow	Good	Fair	Fair	Very Limited	Science
Water Quality	Fair	Fair	Fair	Nil	Science

1. **Fair:** Considerable improvement needed; **Good:** Some improvement needed; **Very Good:** Improvement not needed presently
2. **Very Limited:** A few validation cases; **Partially:** Some cases with successful validation; **Fair:** Considerable validation cases; **Adequate:** Many validation cases
3. **Nil:** No operational application; **Very Limited:** A few well-proven cases; **Some:** Some well-proven cases; **Standard/part.:** Standard professional tool in some regions
4. **Science:** Inadequate scientific basis; **Tech.:** A technological push is required; **Admin.:** Administrative tradition or missing economic motivation

Similarly, we can talk about groundwater pollution or the effect of land use changes. If you talk about the effect of land use changes, say, on flow, their adequacy of scientific basis is good, that is, some improvement is needed. Scientific testing is fair; that is, considerable validation cases are there. And then, if it comes to validation on pilot projects, it's again fair. And when it comes to practical application, this is very limited. So, very limited means a few well-proven cases are there, and when it comes to major constraints for practical application, it's the science.

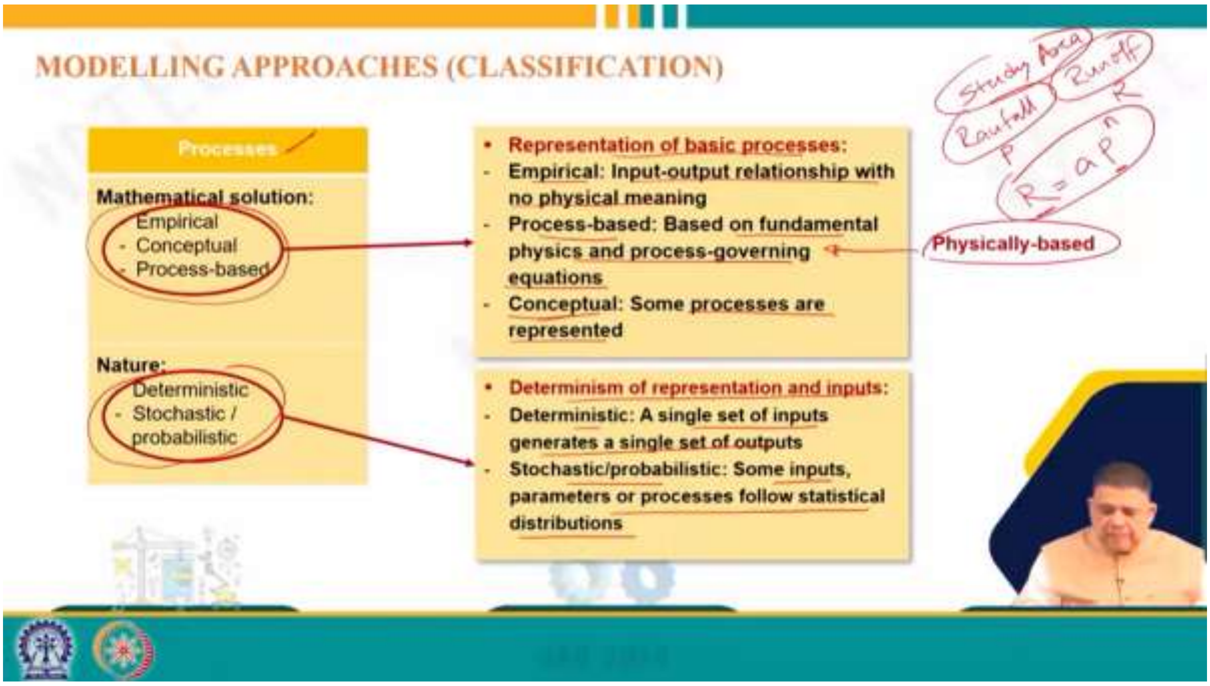
So, science means inadequate scientific basis. So, obviously, there is a lot of scope for improvement in understanding the process itself, and that is why science is. So, obviously, the same is true for water quality; that science is the major constraint. So, obviously, this shows that where science is a constraint, there is a scope for research, where technology is a constraint, there is a technological push required. So, this tells us about the status of modelling applications in general.



Then we come to mathematical models. So, mathematical hydrological models generally have two components: a hydrological core and a technological shell. When we say hydrological core, it consists of the scientific basis for process description. That means you must understand the hydrology well, so that you can get involved or develop the hydrological core. The technological shell consists of programming, user interface, and pre and post-processing facilities. That means how the model interacts or the user interacts with the model, and the background programming, of course. As you can see, as far as the technological shell is concerned, a programmer or a person good in computer programming and developing the interface or maybe graphical user interface can take care of this technological shell. But when it comes to the hydrological core, because it includes the process description and the scientific basis of the process description, only a person who is well aware of the hydrological processes can work.

So, that is why you need a hydrologist to be able to develop the mathematical hydrological models. Mathematical hydrological models are preferred nowadays because of the rapid development with scientific basis and application. So, obviously, as we have seen the status in several areas already, there is good understanding, there are good models, they have been tested, they have been applied in field conditions. That means it is possible to adopt such models and use them in your conditions. And of course, models are universally applicable unless it is an empirical model where directly you are using the data transformation or input-output data relationship.

Other than that, all models where the calibration possibility is there, such models can be universally applied or used. That means, irrespective of whether a model is developed in India, the USA, or somewhere in Europe, you can use that as long as you understand the working of the model in a good way. So, understanding the model is more important, and then you can use any model anywhere unless and until it is an empirical model.



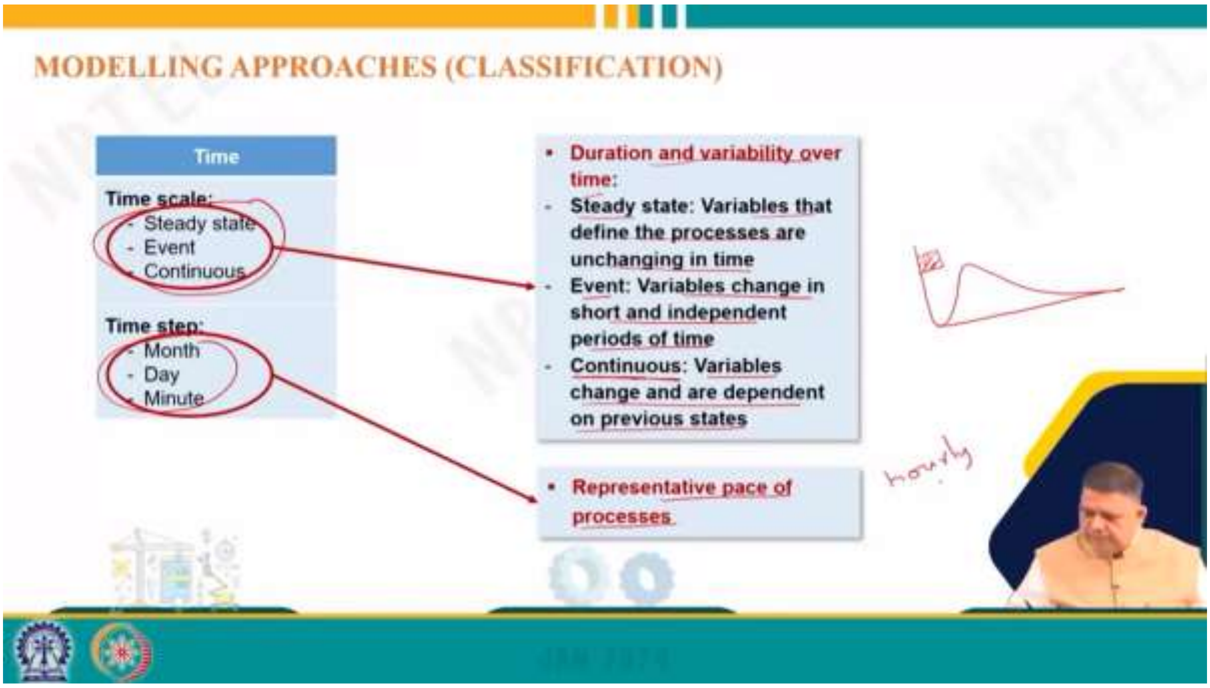
Then we come to the classification of hydrological models. Based on the processes, mathematical models can be classified as empirical, conceptual, or process-based. So, mathematical models are empirical models, conceptual models, or process-based models. It is based on the representation of basic processes. When we say empirical model, it's an input-output relationship with no physical meaning. Basically, if you take a particular study area, you select rainfall data and runoff data. So, rainfall, let's say we call it P, runoff we call R, and then we can develop a relationship between P to the power n or whatever, some relationship we can develop. So, that simply means we have a relationship between rainfall and runoff; runoff is a function of rainfall, and then obviously, the empirical coefficients are there. This is an empirical model and it has no physical meaning, no physical description of the processes are getting involved; you just take the input data, output data, and develop a relationship. That is the empirical model.

On the other extreme, we have process-based models, which are based on fundamental physics and process-governing equations. So, obviously, you take infiltration, you use Richard's equation; you take evapotranspiration, we saw various equations, various hydrological processes, we saw that there are various equations available, some of them are physically based equations based on fundamental physics and governing laws. So, you take those equations and develop that using the model. That is a process-based model, and another term which is used for such kind of models is physically based models, which is a more common term, physically based models.

And in between these two, empirical is just you are not taking any physics into account, anything into account actually, you are just taking input and output data and developing a relationship. In the physically based model, you are taking the physics of the process, that means, in physically based models, you are taking that means, governing laws of mass momentum are coming into picture. So, that is a process-based or physically based model, and then there is a in-between conceptual model where some processes are represented. So, they are a compromise between empirical and process-based. So, some processes are explained, but not in the true sense, maybe some analogy is used which we will see in detail.

So, these are different models: empirical based on process, empirical conceptual, and physically based. Then based on the nature of the mathematical solution process also models can be deterministic, stochastic, or probabilistic. So, when we say determinism, of course, it depends on determinism of representation of input. So, deterministic models a single set of input generates a single set of output so that means, if it is a deterministic model, you run today with the same data or run tomorrow with the same data or Mr. excellence or Miss Y run every time the solution will remain the same.

On the other side, we have stochastic or probabilistic models where some input parameters or processes follow statistical distribution, that means, the probability comes into the picture. So, that simply means when you run such a model that based on the probability that is considered there may be different effects, different results that may come out of such models. So, based on determinism of representation inputs a model could be deterministic or stochastic or probabilistic, that means, the variables which we are considering in the model they could be based on probability or not, I mean whether there is a probability or not. So, if probability is considered, it is a stochastic or probabilistic model; otherwise, it is a deterministic model. Then comes, of course, the model classification based on time scale.

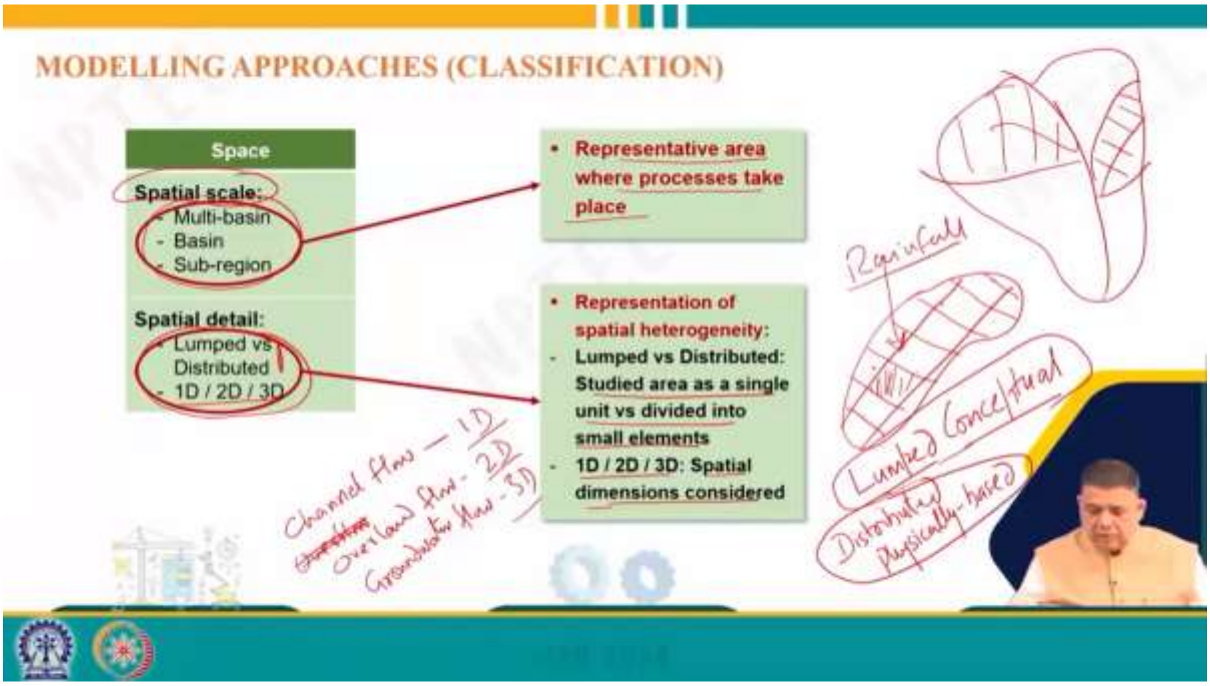


So, a model could be steady-state, event-based, or continuous. Of course, that is duration and variability over time. A steady-state model, as we know, means variables that define the process are unchanging in time, that means the variables will remain constant, the process will more or less remain constant with time, there will be no changes. Then there are event-based models, that is, variable change in a short and independent period of time. Say, we discussed the hydrograph model. So, obviously, there is an event, there is an input event rainfall, and then like you remember, we had a single event and then we plotted the yield hydrograph. So, this is an event-based model, that is, there is a single event and because of that effect, there is a flood hydrograph or a yield hydrograph, whatever you call it.

So, that is an event-based model. Then there is a third type of model, continuous model, which is a more common type of model, continuous where your variable change is dependent on the previous state. So, that means, if you have a model which runs for the entire year, say for example, all 365 days on a daily basis, then obviously, the previous day, what happened the previous day, the soil moisture or the runoff, that also comes into the picture, especially on the input or process modelling side. So, obviously, when you take that into account, the model status of today will vary as per the model status yesterday and so on. So, that is called a continuous model and that means, depending upon whether it is a dry season or a monsoon season, the model behaviour will change. So, that is a kind of continuous model where you model for an entire year or for several years together without stopping in between.

Then, of course, comes the time step. So, the time step use could be a month, day, or minute. So, basically, this represents the pace of processes. So, obviously, if it is a groundwater model where we know that the processes are very slow, we can go for a day or multiple day gap we could have that in the time steps, but if it is something like a hydrograph, then obviously, every minute is important. So, obviously, you will try to have a shorter time span or time step you will use for modelling. So, depending upon that, but hourly or daily is very commonly used time steps.

Then, of course, based on the especially scale also, a model can be classified. That is, a model could be multi-basin, basin, or sub-region, and of course, that represents the area where processes take place. So, obviously, you can have an entire basin model or you can always have a sub-basin model. So, where you have different sub-basins which you try to model. So, depending upon what you are trying to model, a model could be a basin model or a sub-basin model or multi-basin model or sub-region model.

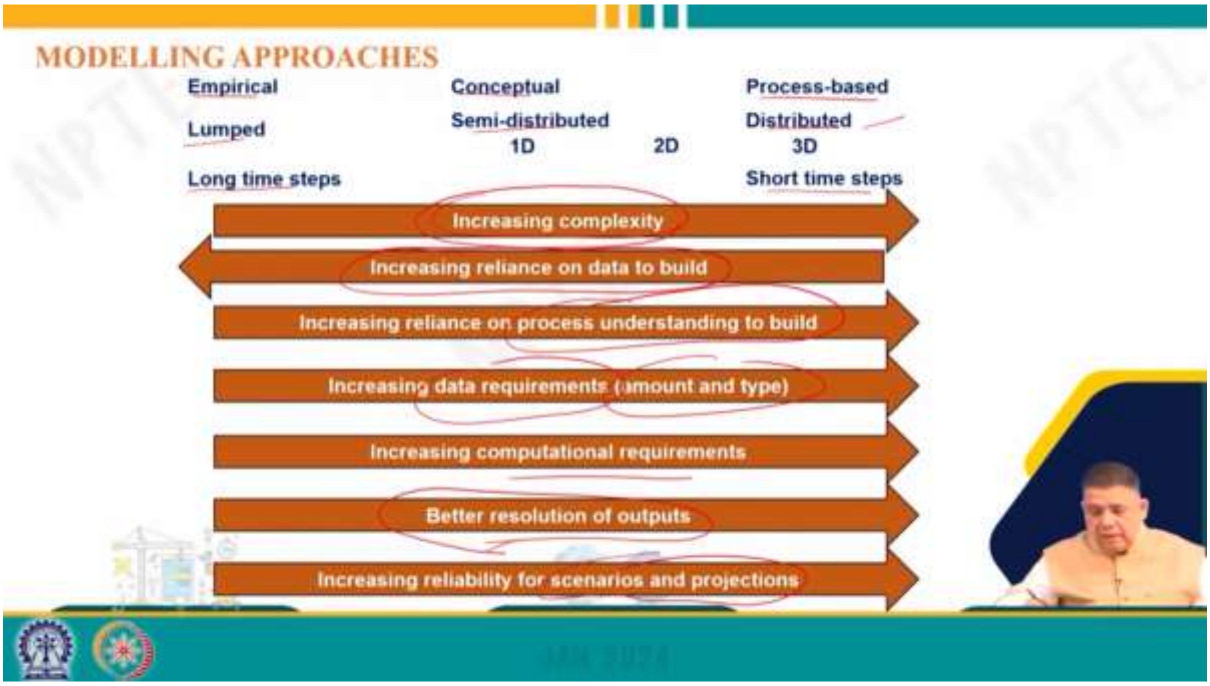


Then, of course, the spatial detail you include in the model, that is, your model could be lumped or distributed. First thing, lumped or distributed, and then obviously, 1D, 2D, and 3D. So, when we say lumped or distributed, that study area is a single unit versus divided into small elements. So, obviously, when you say lumped model, so obviously, what happens is that you consider the entire basin as a single unit. So, obviously, you will have a single rainfall for the entire basin. So, obviously, you remember that even if you have multiple rain gauges, then we will try to use the Thiessen polygon and it method we studied, and we use the mean areal rainfall, mean rainfall as input. So, that is just a single value, single soil characteristics, single everything will be simple.

So, lumped, that is, everything taken together will be used as input and so will be the output. So, that is a lumped model, but if it is a distributed model, so obviously, we can have these models will have the grids of different sizes. So, we can have grids of different sizes, so for different grids, we have different types of inputs and obviously, outputs. So, that is the distributed model, so obviously, if you are using a single value, then it is lumped, and otherwise, it is a distributed model.

And typically, if you talk about the other classification we said, so basically, we have terminology which we use is lumped conceptual model because typically conceptual models are lumped, or we have the other side, we have distributed physically based models.

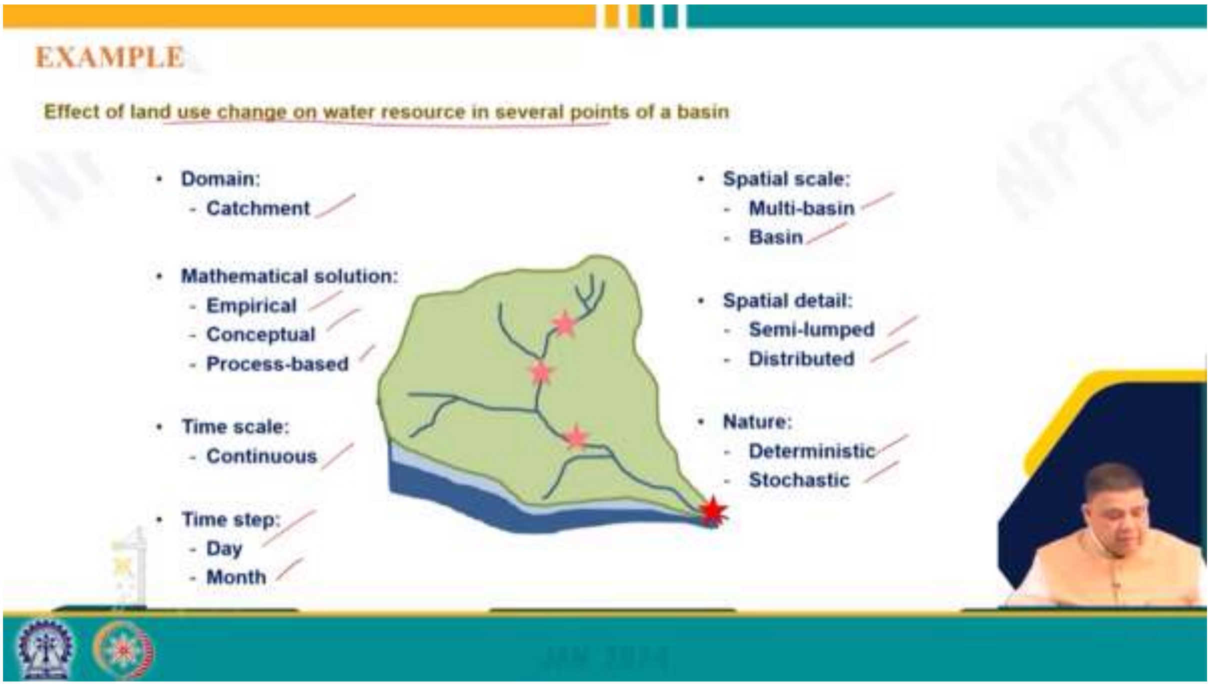
So, typically, physically based models are distributed in nature, whereas our conceptual models will typically be lumped. Of course, there could be distributed conceptual models also, but typically, they will be lumped. So, that is lumping versus distributed, and of course, it could be 1D, 2D, 3D, that is special dimension considered. So, obviously, if it is a channel flow, then obviously, it will be 1D; if it is an overland flow phenomenon, then obviously, we can have 2D; and if it is groundwater flow, then we may consider 3D. So, depending upon the processes, the flow processes, it could be 1D, 2D, and 3D, these are examples. So, based on what you are trying to model and how we are trying to model, based on that, our models will be classified in a different way.



Now, coming to comparing the modelling approaches, if you try to compare the modelling approaches, then obviously, we have seen the classification: empirical, conceptual, process-based; we have said lumped, semi-distributed, distributed; and within distributed, we have also said 1D, 2D, 3D; and we have said long-time steps, short time steps. Now, if you move from left to right on the screen, then obviously, the complexity of the model increases, the model becomes complex. So, an empirical model is simple, as we already saw, you take rainfall runoff, develop a relationship, that is a model. But if you go to a physically based model where you will be describing the physics behind every process, so that is a complex model. So, obviously, if you move from empirical to process-based, complex; if you move from lumped to distributed, because lumped has only a single unit, single input, single output, whereas distributed could have a large number of grids and a large number of inputs. So, that is why the complication is more, and similarly, if it is a long-time step, a monthly model compared to a minute-based model or daily model, the complexity will be more.

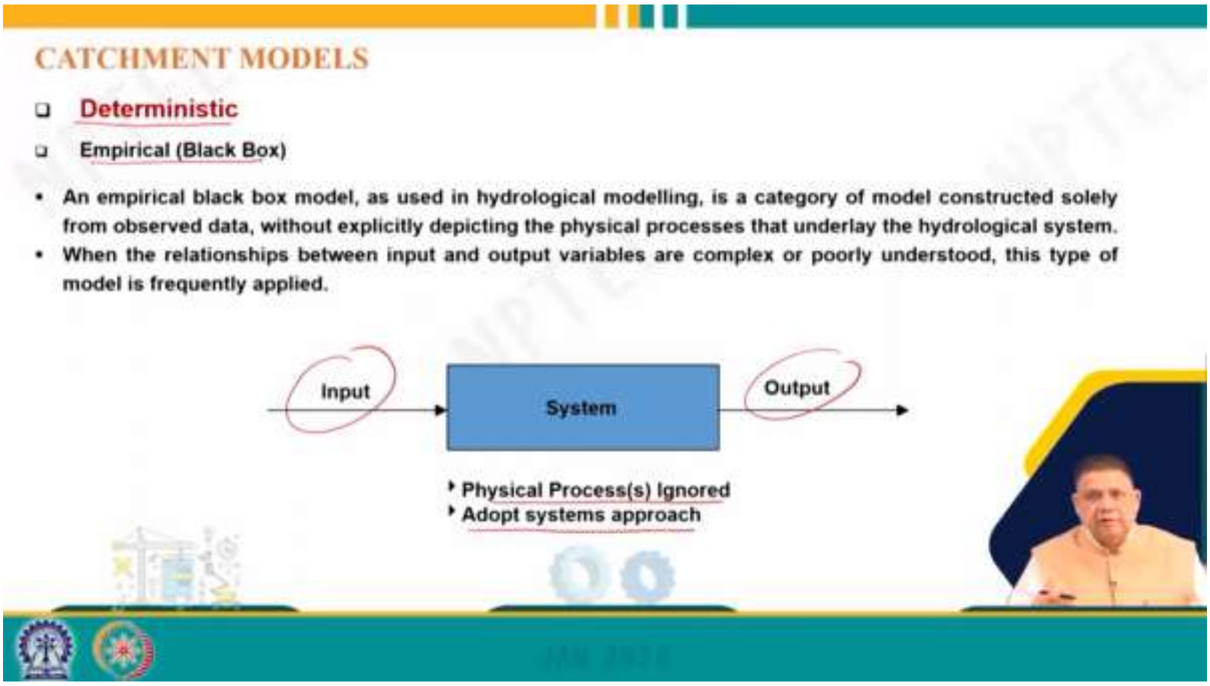
Similarly, if you move from right to left, there will be increasing reliance on data to build. So, obviously, if you want an empirical model, as we saw, we only take the data. So, our reliance is more on the data rather than any understanding. So, if you move from right to left, our reliance on data will increase, increasing reliance on process understanding to build the model.

So, if you move from left to right, then obviously, you have to have a better understanding of the model, and if you move from left to right again, the data requirements will be more, that is, the amount and type of data that is required will be more, which is quite obvious. Increasing computational requirements, better resolution of output, that is the good thing that in an empirical or lumped model, you will get just a single value for the entire watershed, whether on the other hand, if you have a distributed model, then you can have distributed results and an increasing reliability for scenarios and projections. So, obviously, if it is a complex model, then you can have various kinds of scenarios and projections study using this model.



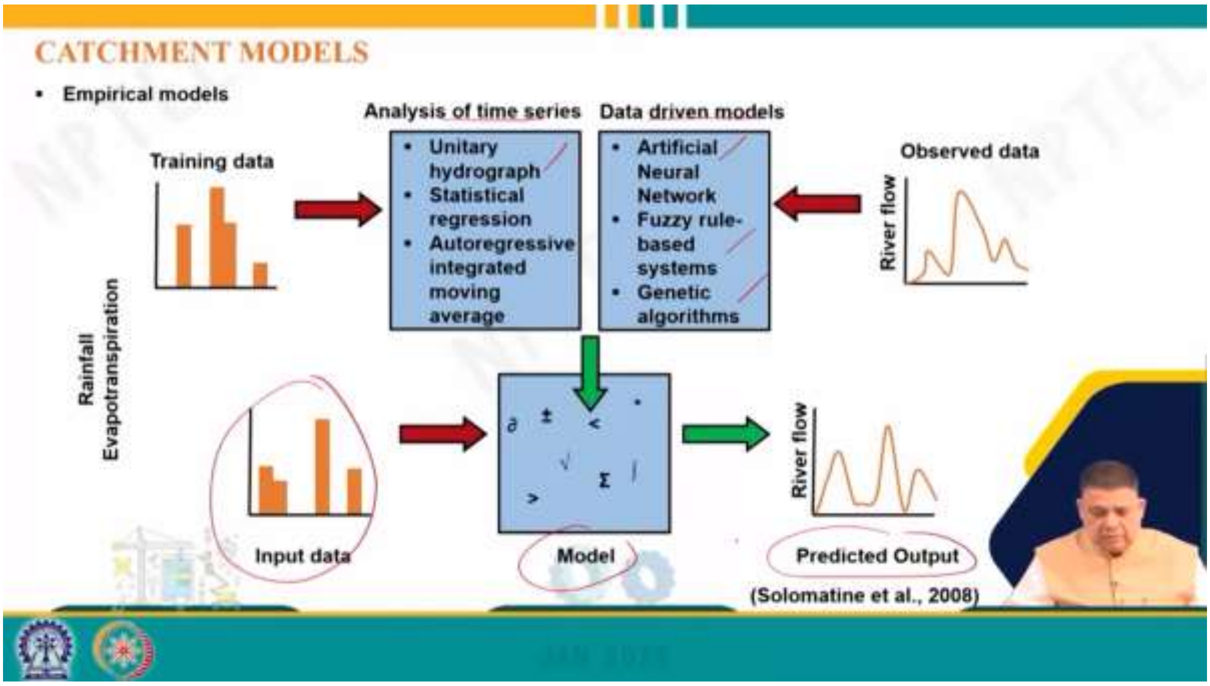
So, this is the modelling approach comparison. If you take an example, basically, then obviously, if you try to study the effect of land use change on water resources, then the domain could be a catchment, then especially scale could be multi-basin or basin.

So, you can only have one measure, one issue, or you can have a multi-basin study. Then mathematical solution, you can use empirical, conceptual, process, anything is possible, you can use semi-lumped or lumped or distributed. Time scale could be continuous, you can have deterministic or stochastic. That means, you can take the variables deterministic or very stochastic and you can adopt a time step daily or even monthly depending upon what you are trying to study. So, that is how you can study various processes, various types of approaches you can adopt while studying a particular analysis.



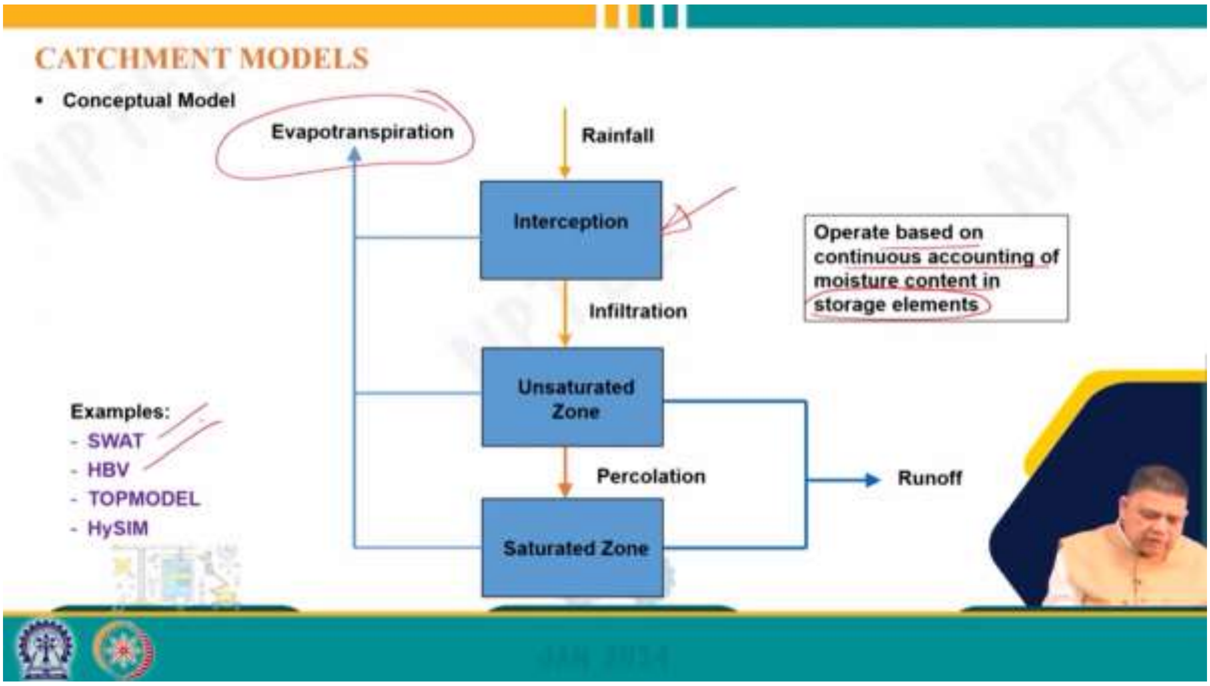
Now, coming to deterministic models, the deterministic model can be further grouped into three categories: empirical black box models, where basically the physical process is completely ignored and it adopts a systems approach. That means, you take the input and just now we saw the example, you take the input, you take the output, and develop a relationship.

So, that kind of model is referred to as an empirical or black box model because the system is completely black; you have no information about the system, you do not know what is happening within the system for converting this input to this output.



So, that is a black box model and typically, all the time series-based models or even hydrograph models or the data-driven models like machine learning, artificial neural network, fuzzy rule-based genetic algorithm, they are all examples of these empirical models where you take the input, there are some mathematical functions, and then you will have the predicted output coming out of it. Then we have the conceptual grey box model, grey box model, which are usually lumped variables and parameters representing average value for the entire catchment, that is what that is basically a spatial description. And though physically sound structure equations are used, the physical significance is not very clear, as we said that we have the empirical and physically based, conceptual is somewhere in between.

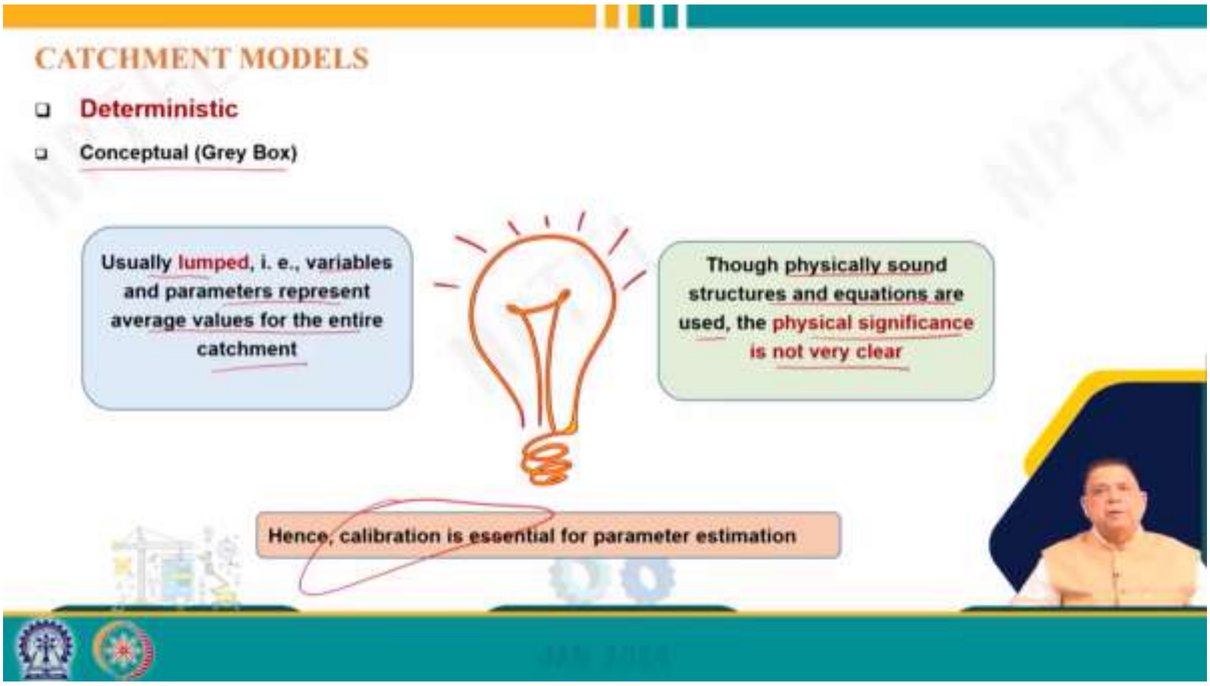
So, though the physically sound structure and equations are used, but physical significance is not clear and that is why calibration is essential for parameter estimation of these models.



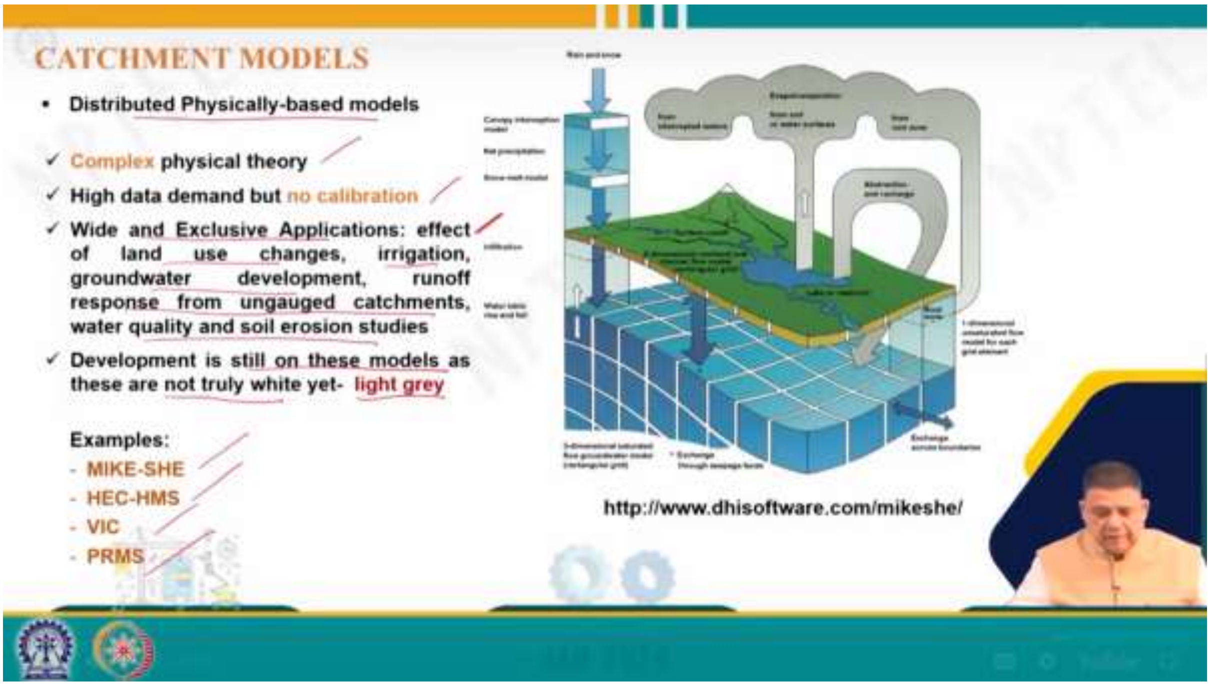
So, for example, typically, if you are conceptualizing a rainfall-runoff model, then you can take the help of storage elements. So, a two-box model, three-box model. So, here we can say that my top box represents interception and it can have the assumption that rainfall occurs to satisfy the interception, only when interception is satisfied, water will infiltrate and go to the unsaturated zone and so on and so forth.

So, that kind of assumption you can make. So, basically, such models operate based on continuous accounting of moisture-containing storage elements. So, obviously, when such unsaturated zone is satisfied, then percolation will occur and infiltration rate is high, then runoff will be generated. And evapotranspiration could take place from interception as well as from the unsaturated zone or saturated zone, the rate could differ depending upon how you are formulating the model.

So, basically, you have the physical processes and they are equations, but the representation is not true, it is more simplified, that is why it is a conceptual model. There are several models like that, for example, SWAT, HVV, TOP model, HI-SIM, these are all examples of conceptual models.

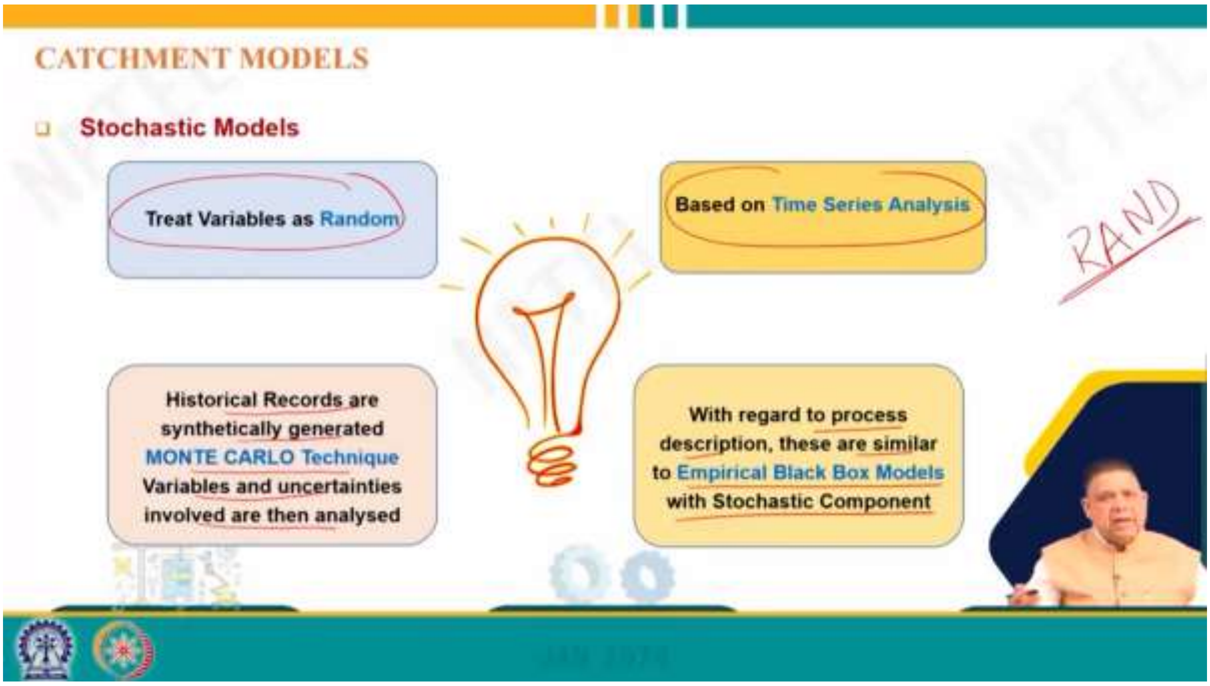


Then we have the third category, physically based white box models, which describe the natural system using equations of mass, momentum, and energy, and parameters have direct physical significance. Usually, these models are fully distributed. Of course, because these models have direct physical significance, calibration is not required, that means, a length less length of data could be required and you can use. So, the hydrological fraternity is divided into separate schools, conceptual versus physically based, they are two schools of thought, some say that conceptual is better or good enough, some say physically based is good enough.



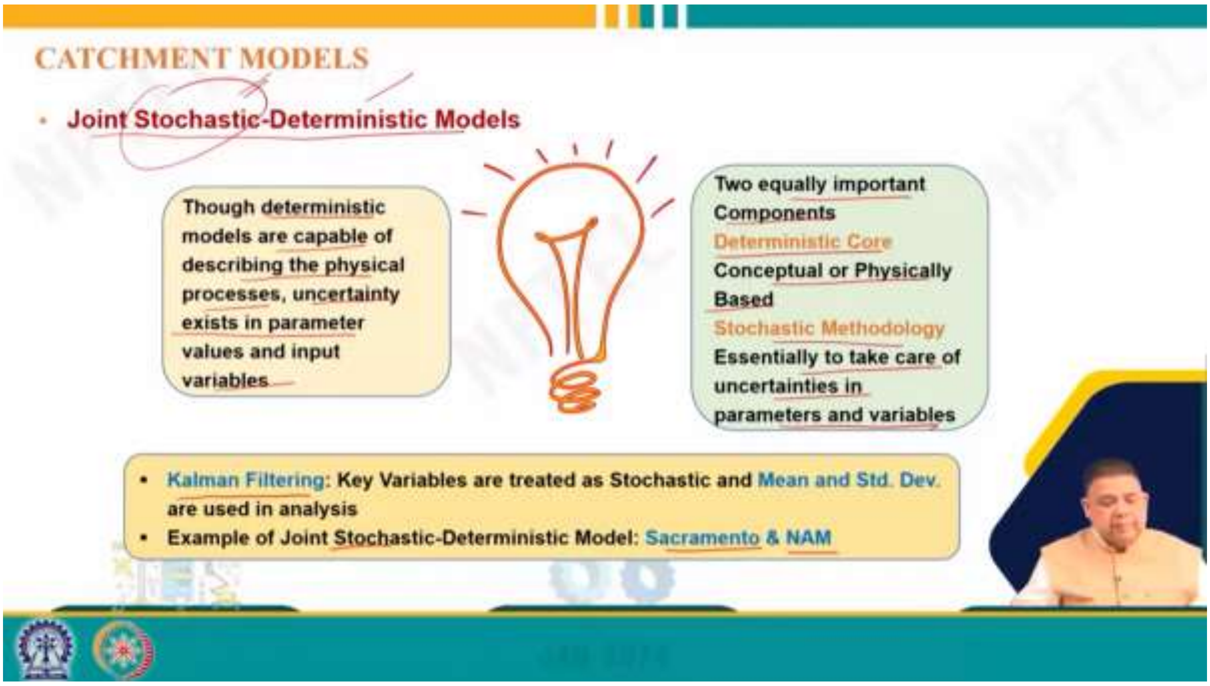
So, that is why, though distributional physical-based models, they are not yet truly development is still on and these are not truly white yet, they are light Gray, but there are some exclusive applications of such models.

So, they are they have complex theory, high data demand, but they have wide and exclusive applications. For example, the effect of land use changes, irrigation, groundwater development, runoff, rain form from ungagged capsids, water quality, and soil is used to study all those can be done only using a physically based model that is not possible with a conceptual model or an empirical model. And examples of physically based models are Mike sheen, HEC-HMS, WIC, PRMS. So, these are some of the models and with time you will study about these models in some other courses at a higher level maybe. Then we have stochastic models, we have already seen deterministic and stochastic.



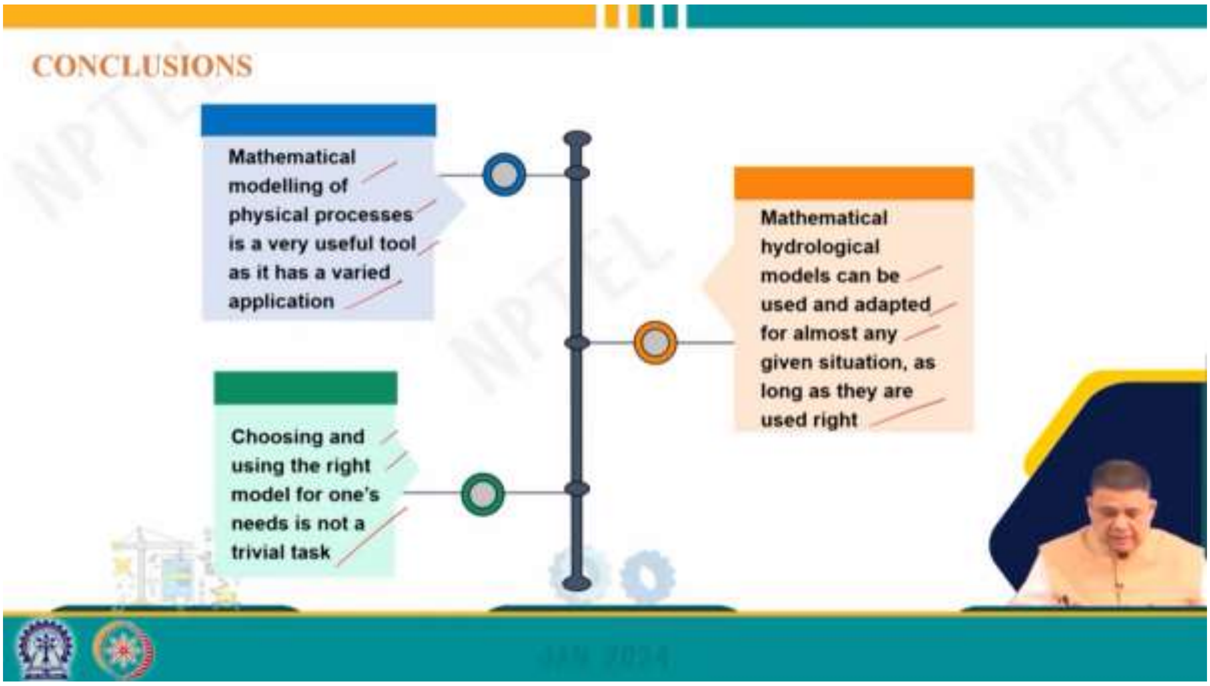
So, deterministic we just saw, now coming to stochastic, they treat variables as random and basically, they are based on time series analysis. So, historical records are synthetically generated, you will see say Monte Carlo technique and variables are certainly all then and light with regards to process description, these are similar to empirical black models with stochastic components. So, there is no physical process involved in these models. So, basically, they work on the random number generation, the Monte Carlo technique is one you can even use Excel, Excel there is a function called rand. So, using that rand function, you can generate the data, random number generation.

So, you take a parameter you know the range and then you can generate hundreds of numbers. So, using the rand function you can do that and then you use the mean and standard deviation and then use that into modelling to take care of some uncertainty. And of course, the deterministic and stochastic model, this we already saw: in a deterministic system, a single cause will always give one single effect, but in a stochastic system, a single cause may give you different effects depending upon the probability that is being considered.



Then of course, nowadays there's a lot of talk about joint stochastic deterministic models because the deterministic and stochastic models have their own advantages and disadvantages. Deterministic models are capable of describing the physical processes, but they have uncertainty in parameters and input variables. Wherein stochastic models, because they take the variables as random and they take the mean and standard deviation through random number generation, so obviously, uncertainty is taken care of. So, that is why joint stochastic models have two equally important components: a deterministic core that is conceptual or physically based core will be there, but methodology will be stochastic, that is essentially you take care of uncertainty in parameters and variables. So, that is how you take advantage of both deterministic and stochastic models and that is why joint stochastic deterministic models are becoming popular. Kalman filtering is one important way of handling joint stochastic models and examples are Sacramento and Nam, these are the two models which are examples of joint stochastic models.

So, coming to a few popular models, as you can see that as I already mentioned there are hundreds of thousands of models available and as you can see that there are models popping up there on the screen and of course, the models could be very complicated, say Mike Sheen, that is a physically based distributed model, they can be simple ones like HVB, is a relatively simple model, then there were Boston basin models and so on so forth. Various models are available, but obviously, while using, you have to be careful about and you have to have knowledge about the assumptions behind those models and the purpose for which they have been developed.



So, with this, we conclude by saying that mathematical modelling of physical processes is a very useful tool, and it has varied applications, as we have seen throughout mathematical hydrological models can be used and adopted for almost any given situation as long as they are used right. So, obviously, you have to know where and how and then you can use the models effectively, and choosing and using the right models for one's need is not a trivial task. So, though we may say that modelling is very, I mean, there are misconceptions, but choosing the right model and using it the right way is not a trivial task, it requires considerable knowledge.

So, with this, we close this topic. Thank you very much. Please give your feedback and raise your concerns or doubts, and we will be happy to answer on the forum. Thank you very much.