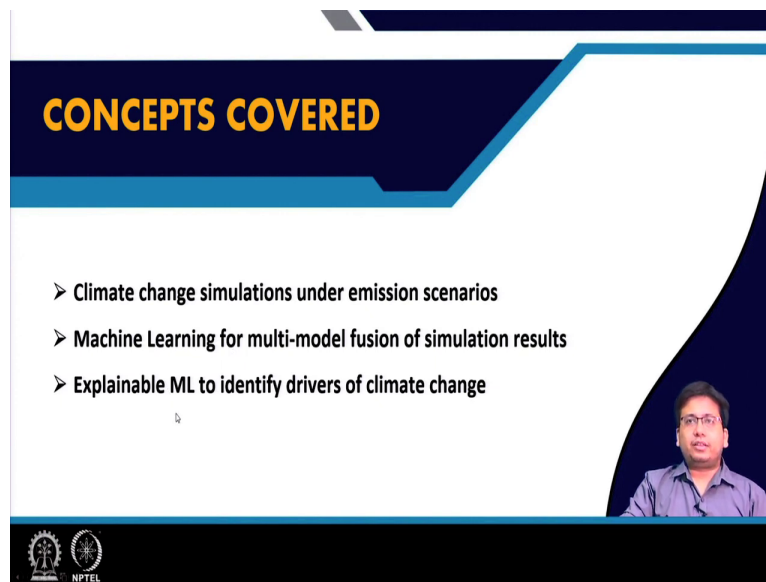


**Machine Learning for Earth System Sciences**  
**Prof. Adway Mitra**  
**Department of Computer Science and Engineering**  
**Centre of Excellence in Artificial Intelligence**  
**Indian Institute of Technology, Kharagpur**

**Module - 05**  
**Machine Learning for Earth System Modelling**  
**Lecture - 40**  
**ML for Climate Change Projection and Course Conclusion**

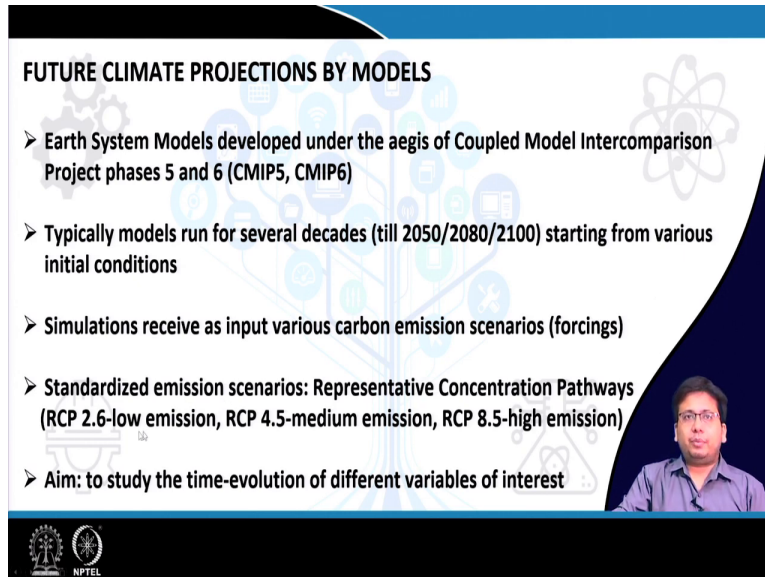
Hello everyone. Welcome to the 40th and last lecture of this course on Machine Learning for Earth System Science. This is the last lecture of module 5 which is related to Machine Learning for Earth System Modelling and the topic of this lecture is going to be Machine Learning for Climate Change Projection and after this, we will just do a brief review of the course because this is the last lecture.

(Refer Slide Time: 00:51)



So, the concepts which we are going to cover today are climate change simulations under different emission scenarios; how machine learning can be used to that is to a fusion of the different models which simulate this climate change under different scenarios and also, how explainable machine learning can be used to identify the different drivers of climate change.

(Refer Slide Time: 01:16)



**FUTURE CLIMATE PROJECTIONS BY MODELS**

- Earth System Models developed under the aegis of Coupled Model Intercomparison Project phases 5 and 6 (CMIP5, CMIP6)
- Typically models run for several decades (till 2050/2080/2100) starting from various initial conditions
- Simulations receive as input various carbon emission scenarios (forcings)
- Standardized emission scenarios: Representative Concentration Pathways (RCP 2.6-low emission, RCP 4.5-medium emission, RCP 8.5-high emission)
- Aim: to study the time-evolution of different variables of interest

The slide features a background with a blue and white geometric pattern. A small video inset in the bottom right corner shows a man with glasses and a blue shirt. The NPTEL logo is visible in the bottom left corner.

So, first of all, the like we all know that climate change is a very real threat to the earth and human civilization in the coming decades. And so, it is very important for scientists and policy makers to understand what the different climate scenarios are going to look like in like a few decades into the future. And so, for that various earth system models have been developed under the like under the aegis of this coupled model inter comparison project. So, this CMIP5.

So, this CMIP5 is basically, it is a project to compare the climate models being developed by the different research agencies all over the world. So, like CMIP6 is the latest version of the of these models. Before this, there was CMIP5, the fifth version and so on. Now, typically these kinds of models these are all earth system models which we have like GCMs which we have already discussed earlier several times.

So, these models are to be what happens in this simulation is the simulation studies is these models are run for several decades into the future, starting from various initial conditions and they receive as input different carbon emission scenarios. So, the that is we understand that climate change is a function of the various activities which we as a human civilization undertake.

So, there are various carbon emission scenarios. Carbon is in this case, the carbon emission is considered as the most important drivers of the climate change and what the result will be or how

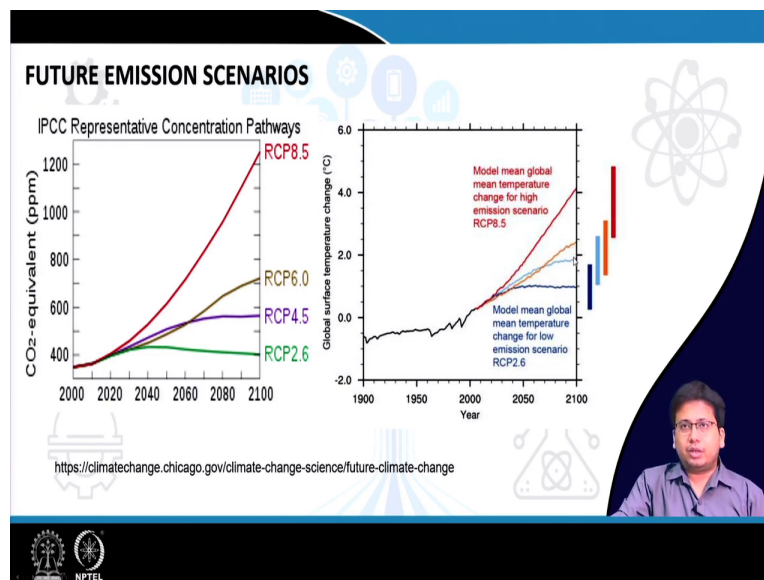
the climate will be in the future largely depends on how we manage our carbon emissions in over the time.

So, for the now, this carbon emissions, this may this is like this is of course non-homogeneous. It is not homogeneous either over space or what time that is we know that there are many countries which emit a lot of carbon or even within the same country, there might be several regions which are emitting more carbon; other regions which are emitting less and so on.

But like taking everything into account, if we like the scientists have identified certain prototype pathways of that is of emission over the years. So, these are some these are like basically they are called as RCP 2.6, this means there is there will be overall low emission of carbon or RCP 4.5 means medium emission; RCP 8.5 is high carbon emission etcetera.

So, like here these 8.5, 4.5, 2.6, these are actually some measures of emission in watts per square meter. The aim here is to study the time evolution of the different variables of interest under all these emission scenarios that is if we are emitting low carbon, then how will the temperature vary over time and such things.

(Refer Slide Time: 04:23)



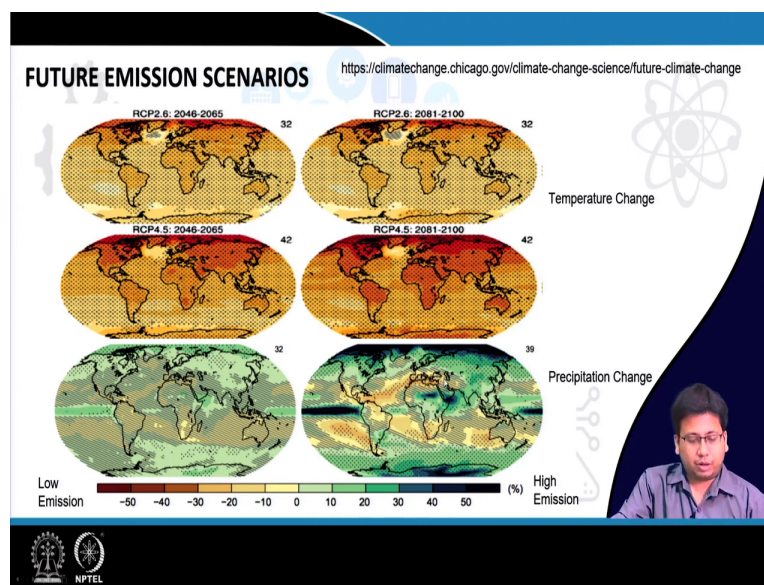
So, like under the different scenarios like if we plot the total CO<sub>2</sub> in the atmosphere, how the total I mean the how the volume of total CO<sub>2</sub> varies over the years. Then, we see a huge

difference. If you it if you are going for RCP 2.6, then we see the total like you know currently, we are somewhere here. If we are going for RCP 2.6, we see that the total carbon in the atmosphere roughly stabilizes.

But in RCP 4.5 or RCP 6.0, we find that they are they continue to steadily increase that is and as a result of that global temperature also gradually increases, while in this case RCP 8.5, it just goes up and goes beyond control. And so, like if this is the scenario, then the basically it means that the earth will become an unlivable place by 2100. But if we go for RCP 2.6, the earth may still be little bit hotter than it is now; but it will like, but it will stabilize; there will not be any further problems.

In the intermediate scenarios, like it may still be somewhat under control; but the problems will keep on multiplying with every given year as we are already seeing now that is due to frequent heat waves and so on.

(Refer Slide Time: 05:40)



So, in the different future emission scenarios, so like these are how the temperature might change. So, in RCP 2.6 which is a low emission and this is how the temperature might evolve during the like period 2046 to 2065 compared to what it is currently. So, while in case of RCP 2046 to RCP 2065. In this case like this is what the like it will be. So, in this like.



So, do not go for this axis for the for this color axis for the first two images. So, like it shows that in these like the temperature will gradually increase in like in both scenarios RCP 2.6 as well as RCP 4.5, but the. So, like here the what is being shown here is the amount of increase in temperature in these periods compared to the current periods and here, you can see that like in the RCP 2.6, the heating in around 2100, I mean at the end of the century will not be very different from what it will be at around roughly around 2050.

That is 2050 onwards, there will not be any very significant heating. The same cannot be said at RCP 4.5; here around 2050, we will see a considerable heating compared to current time. And like at the end of the century, we will see even further heating. While this is regarding temperature; but while we come to rainfall, we see that in like in a low resolution setting, I am sorry in a low emission setting, we may see like decrease of rainfall in some places and increase of rainfall in some other places as indicated by this shading.

So, the shade the brownish means reduction of rainfall and greenish or bluish means increase of rainfall. So, as we can see there is going to be reduction of significant reduction of rainfall in the Southern hemisphere; but significant increase of precipitation in the Northern hemisphere; especially in this Russian and North European region.

While these tendencies we, but these are all light tendencies in the low emission scenario; but in the high emission scenario, both this reduction and the increase, both of them will be much more pronounced. That is the Southern hemisphere will receive very less rainfall, while the northern hemisphere may be receiving much higher rainfall.

(Refer Slide Time: 08:36)

**Issues with Future Scenario Analysis**

- Huge uncertainty across models
- Different models accurate for only specific variables and/or specific regions even in historical settings
- Climate models are computationally extremely expensive
- Anthropogenic forcings hard to predict (shared socio-economic pathways try to)
- Predicting earth system response to forcings difficult due to feedback among different variables
- Identifying high-impact extreme events difficult

The slide features a background with various icons representing technology and science. A small video feed of a male presenter is visible in the bottom right corner. The NPTEL logo is in the bottom left corner.

Now, there are various issues with this kind of future scenario analysis. So, first of all, there is huge uncertainty across the models. As already mentioned, there are different research groups that come up with different models and each of these models simulate the future scenarios under like under the I mean the future climate under the different scenarios. But there is significant disagreement among the models in this matter.

So, for example, if you see here what the global surface temperature change will be in the different models model simulations; you can see these bars here. So, these bars are basically the spread as observed in the simulations by the different models. So, some models may be predicting that the temperature rise will be something like this, while the other say that the temperature rise might be something like this which is a significantly high.

In this case also we can see this kind of a spread. So, these like these uncertainty across models is a big problem in like making any kind of prediction about the future scenario. Apart from that, there are different models are that is the thing is that we cannot trust any of these models. That is there is no single model which we can say is the correct model because it is found that different models accurate only like are they are accurate for only specific variables or specific regions, even in the historical settings.

That is forget the future, even the period for which we have observations, there is no model which is able which is able to simulate the this the historical period also with perfect accuracy. Some are perfect in some regions, others are perfect in other regions and so on. Also, another big problem is that these climate models, these are extremely expensive to for as far as computation is concerned.

That is they require supercomputers to be running and then, also another problem is that the anthropogenic forcings are hard to predict that is we can cannot be sure about how exactly the emissions will be panning out; I mean we like of the different scenarios like 2.6, 4.5 etcetera have been discussed.

But even like each of these scenarios also like is like can be divided into various sub scenarios in which some parts of the world emit more, some parts of the world emit less and these kinds of things. Now, again which part emit which regions emit more, which regions emit less that also have will have a bearing on how exactly the climate change will pan out.

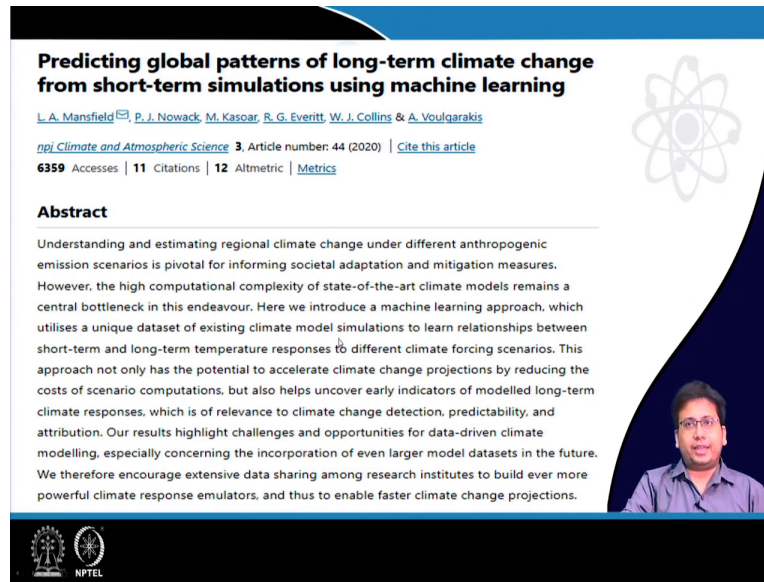
But these kinds of things are more difficult to predict and so, as a result of that what the earth systems response will be to all these kinds of forcing is difficult. Another source is the feedback among the different variables that is if one variable keeps rising very much, it may trigger of and the rise of another variable which may impede the first variable or which may actually catalyze the variable to rise further like.

So, these kinds of feedback are also there and they are not under often not understood very well. So, as a result, there is certain amount of uncertainty is also present in these models because of our inadequate understanding of the physics.

And another is that like the what really matters is how many extreme like events take place that have very high impacts that is how many cyclones are expected, how many heat waves are expected, how many droughts are expected because these are the things which really like hamper human life rather than the average temperature itself rising, it is really these extreme events which impact cause more impact.

But like estimating them, according from simulations is not very straight forward. Because the simulations are aimed to simulate specific variables. They are not aimed to simulate likes particular extreme events like this.

(Refer Slide Time: 12:42)



**Predicting global patterns of long-term climate change from short-term simulations using machine learning**

[L. A. Mansfield](#) , [P. J. Nowack](#), [M. Kasoar](#), [R. G. Everitt](#), [W. J. Collins](#) & [A. Voulgarakis](#)

*npl Climate and Atmospheric Science* **3**, Article number: 44 (2020) | [Cite this article](#)

**6359** Accesses | **11** Citations | **12** Altmetric | [Metrics](#)

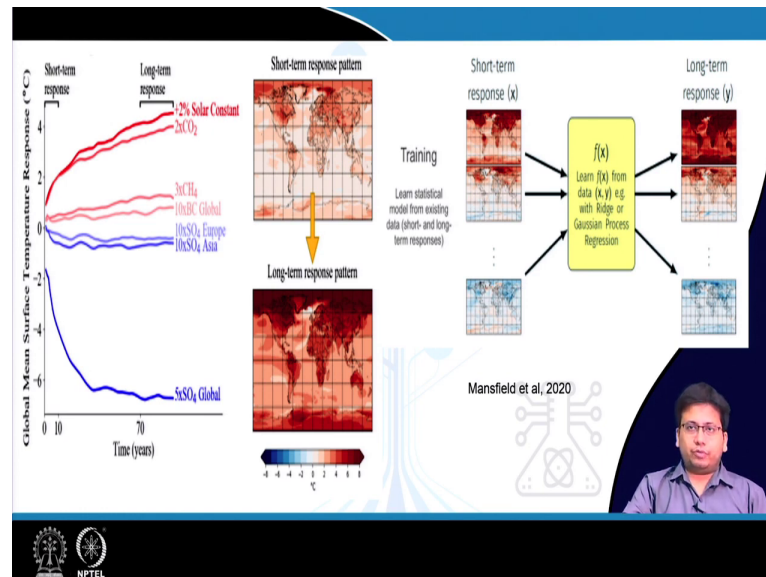
**Abstract**

Understanding and estimating regional climate change under different anthropogenic emission scenarios is pivotal for informing societal adaptation and mitigation measures. However, the high computational complexity of state-of-the-art climate models remains a central bottleneck in this endeavour. Here we introduce a machine learning approach, which utilises a unique dataset of existing climate model simulations to learn relationships between short-term and long-term temperature responses to different climate forcing scenarios. This approach not only has the potential to accelerate climate change projections by reducing the costs of scenario computations, but also helps uncover early indicators of modelled long-term climate responses, which is of relevance to climate change detection, predictability, and attribution. Our results highlight challenges and opportunities for data-driven climate modelling, especially concerning the incorporation of even larger model datasets in the future. We therefore encourage extensive data sharing among research institutes to build ever more powerful climate response emulators, and thus to enable faster climate change projections.



And now, this paper which appeared in nature communication sorry in npa journal for climate and atmospheric sciences. So, the aim of this paper is to solve one of the problems which we discussed earlier; namely, that the climate models are computationally very expensive.

(Refer Slide Time: 13:30)



Here, like if suppose I want to know what the what is the least this model is going to predict for 2100 or 2080, this period which is far in the future, the thing is like I should idea like in the base case, we will have to run the model till that point of time and to know what it is going to simulate. But what this paper is trying to say is that you do not really need to do that; you for most models by looking at its short term responses, you might be able to predict its long term response also.

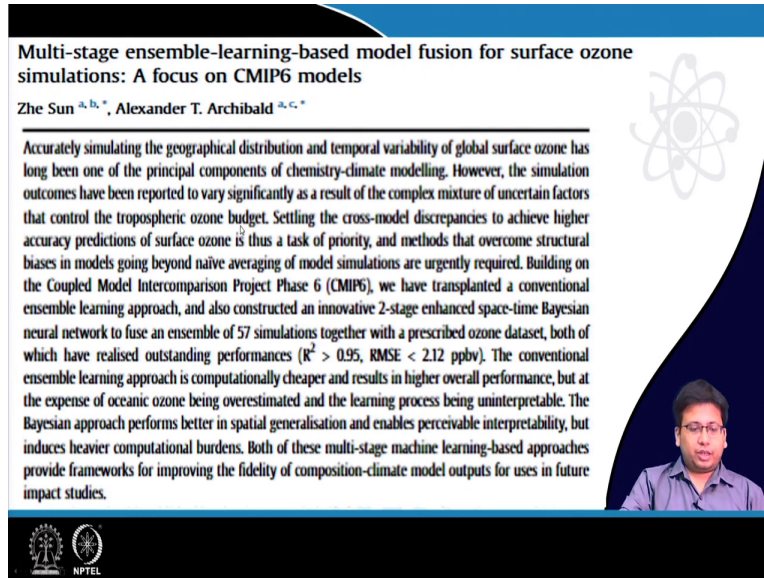
So, suppose I run the model I want to know the what will be the response of the model 70 years or 100 years from now. So, what I will do is I will run the model only for the next 10 years which will be of course less; I mean which will still require computation. But certainly less than 100 years and, but based on what it is; what it is going to predict for the next 10 years, I will understand what it is going to predict after 100 years also. That is its short term response will have to be mapped to the long term response and so that is an interesting idea.

Because like that basically means that I do not really need to run the model for the entire period, just based on a small period, I am able to predict what it is how much it will predict into the or what it will predict if it is left to run in the till the future also. So, they basically train a machine learning model which maps the short term response of the model; I mean the by short term

response, I mean how the variables are going to evolve 10 years from now starting from a particular initial conditions.

So, that short term response is mapped to the long term response like this. So, they have considered Gaussian process regression and so on as models.

(Refer Slide Time: 15:15)



**Multi-stage ensemble-learning-based model fusion for surface ozone simulations: A focus on CMIP6 models**

Zhe Sun <sup>a,b,\*</sup>, Alexander T. Archibald <sup>a,c,\*</sup>

Accurately simulating the geographical distribution and temporal variability of global surface ozone has long been one of the principal components of chemistry-climate modelling. However, the simulation outcomes have been reported to vary significantly as a result of the complex mixture of uncertain factors that control the tropospheric ozone budget. Settling the cross-model discrepancies to achieve higher accuracy predictions of surface ozone is thus a task of priority, and methods that overcome structural biases in models going beyond naïve averaging of model simulations are urgently required. Building on the Coupled Model Intercomparison Project Phase 6 (CMIP6), we have transplanted a conventional ensemble learning approach, and also constructed an innovative 2-stage enhanced space-time Bayesian neural network to fuse an ensemble of 57 simulations together with a prescribed ozone dataset, both of which have realised outstanding performances ( $R^2 > 0.95$ ,  $RMSE < 2.12$  ppbv). The conventional ensemble learning approach is computationally cheaper and results in higher overall performance, but at the expense of oceanic ozone being overestimated and the learning process being uninterpretable. The Bayesian approach performs better in spatial generalisation and enables perceivable interpretability, but induces heavier computational burdens. Both of these multi-stage machine learning-based approaches provide frameworks for improving the fidelity of composition-climate model outputs for uses in future impact studies.



As now, we come to another paper in this like this is more to do with this problem of having huge uncertainties across the models. So, we know that there are different models which simulate in different ways. So, accurately, simulating the geographical distribution and temporal variability of global surface ozone has long been one of the principal components of chemistry climate modeling. So, here, they are focusing on one particular variable; namely, ozone, the surface ozone.

However, the simulation outcomes have been reported to vary significantly as a result of complex mixture of uncertain factors that control the tropospheric ozone budget. Settling the cross-model discrepancies to achieve higher accuracy predictions of surface ozone is thus a task of priority, and methods that overcome structural biases in models going beyond naïve averaging of model simulations are urgently required. So, one possible like approach is of course you have the simulations by all the models, you just take an average of them.

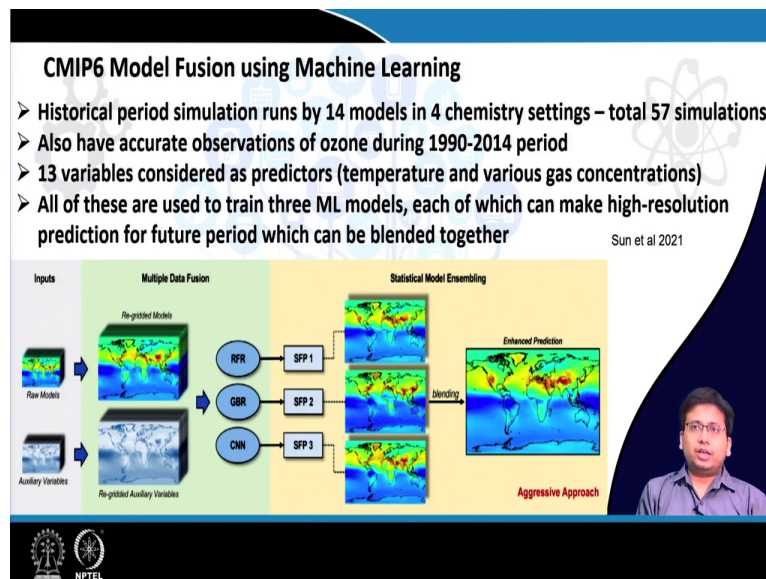


But that is hardly ideal because some as already mentioned, some of the models might be skillful in some regions, some models may be more skillful in other regions and so on. So, like it is necessary that at every location, we weight the simulate or we attach weights to the simulations by the different models and these weights will be varying from one region to another may be one time to time and so on.

So, like building on the CMIP6, we have transplanted a conventional ensemble learning approach and also constructed an innovative 2-stage enhanced space-time Bayesian neural network to fuse an ensemble of 57 simulations together with a prescribed ozone data set, both of which have realized outstanding performances. The conventional ensemble learning approach is computationally cheaper and results in higher overall performance, but at the expense of oceanic ozone being over estimated and the learning process being uninterruptible.

The Bayesian approach performs better in spatial generalization and enables perceivable interpretability, but induces heavier computational burdens.

(Refer Slide Time: 17:38)



So, these are the two approaches which they are talking about. So, like in this. So, basically, they have like total 57 simulations by 14 models and apart from, so like let us say that these are the raw model simulations. Like what is being shown here is the mapping of ozone across the world,

surface ozone across the world. They also have accurate observations of ozone in this period 1990 to 2014. Apart from that, they have 13 variables which can be considered as predictors of surface ozone; namely, the temperature and different gas concentrations.

Now, all of these are used to train the three machine learning models. So, they have considered random forest, gradient boosted regression and convolutional neural network. So, like each of these models are trained separately like using the; like using the these data that is like they have the simulations for surface ozone and they also have the these the actual observations.

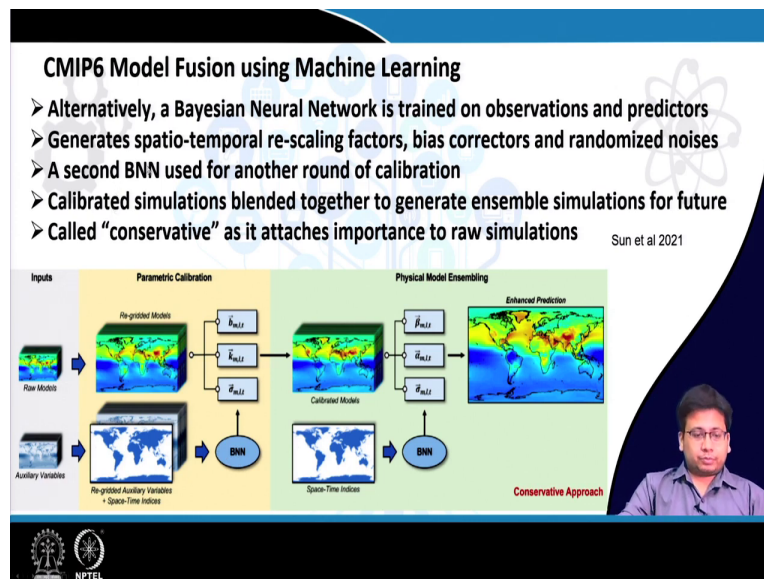
So, based on that in the historical period, they train three different models which we will be able to map; each the simulations by each model to the ground truth. So, they also apart from that, they also have apart from these like model simulations they also have these predictor variables. So, at each model, we will be able to map a given model simulation as well as the predictor variables to the actual quantities of ozone as observed.

Now, once these models have been trained, these models will be like deployed in the future scenarios for which there are no observations. So, the simulations by the different models will then be like provided as inputs to the trained neural networks. So, each of them will produce corrected version of these of the model simulation.

Because the these models have already been trained to predict the correct or the accurate observations of zones like the or by taking into account the simulations and so, like from each set of models, we will have the corrected like the corrected simulations and then, all of them will somehow be blended together by averaging or so on and then, we will get what is known as the enhanced prediction.

So, like in a sense, it is like taking the multimodal ensemble; but like with an additional step of this machine learning models. That is the simulations from the different models are first provided into the three different machine learning models, the models predict some corrected versions of those simulations and then, those corrected versions are blended together to produce like a super corrected version of the enhanced version. This is one approach.

(Refer Slide Time: 20:39)



The other approach is to develop a Bayesian neural network, which is like trained on all the observations and the predictors. So, this is the Bayesian neural network. So, what the Bayesian neural network learns is it manages to generate some spatio-temporal re-scaling factors, bias corrections and randomized noise. That is for each of the locations it should be able to predict some kind of like bias and some kind of rescaling factors and so on.

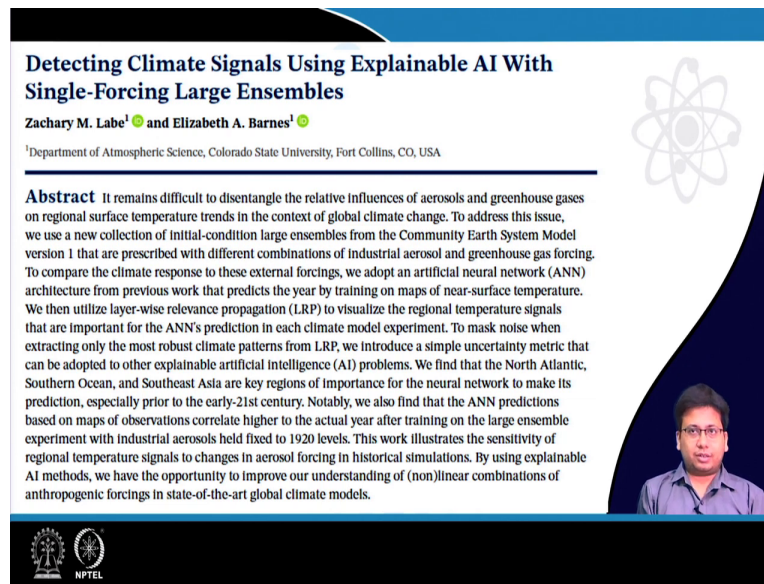
And like another round of calibration will also take place based on something known as something which they call as space time indices. So, based on that the space time indices as well as the these all those predictor variables etcetera, this first BNN is trained. The second BNN is trained only based on the these space time indices.

So, basically what the in the in case of the Bayesian neural network, what is what is found out is like the different biases and other rescaling factors for at for each location and each point of time and so on. So, based on that, all the simulations by the different 57 models as mentioned, like they are somehow re-scaled, re-weighted, bias corrected etcetera to get the final enhanced prediction.

So, as they mentioned the like the first approach using like machine learning to predict the errors, these results in like this is computationally cheap and the overall accuracy is higher. But in

certain specific regions, there is over estimation; in certain other regions, there is under estimation. In this case, those effects are reduced, even though the overall error might be slightly higher and the computationally also this approach is expensive.

(Refer Slide Time: 22:34)



**Detecting Climate Signals Using Explainable AI With Single-Forcing Large Ensembles**

Zachary M. Labe<sup>1</sup> and Elizabeth A. Barnes<sup>1</sup>

<sup>1</sup>Department of Atmospheric Science, Colorado State University, Fort Collins, CO, USA

**Abstract** It remains difficult to disentangle the relative influences of aerosols and greenhouse gases on regional surface temperature trends in the context of global climate change. To address this issue, we use a new collection of initial-condition large ensembles from the Community Earth System Model version 1 that are prescribed with different combinations of industrial aerosol and greenhouse gas forcing. To compare the climate response to these external forcings, we adopt an artificial neural network (ANN) architecture from previous work that predicts the year by training on maps of near-surface temperature. We then utilize layer-wise relevance propagation (LRP) to visualize the regional temperature signals that are important for the ANN's prediction in each climate model experiment. To mask noise when extracting only the most robust climate patterns from LRP, we introduce a simple uncertainty metric that can be adopted to other explainable artificial intelligence (AI) problems. We find that the North Atlantic, Southern Ocean, and Southeast Asia are key regions of importance for the neural network to make its prediction, especially prior to the early-21st century. Notably, we also find that the ANN predictions based on maps of observations correlate higher to the actual year after training on the large ensemble experiment with industrial aerosols held fixed to 1920 levels. This work illustrates the sensitivity of regional temperature signals to changes in aerosol forcing in historical simulations. By using explainable AI methods, we have the opportunity to improve our understanding of (non)linear combinations of anthropogenic forcings in state-of-the-art global climate models.

Now, comes like and in this paper, they use the explainable AI concepts like layer-wise relevant propagation which we have already discussed earlier to identify certain important drivers of climate change. So, it remains a difficult to disentangle the relative influences of aerosols and greenhouse gases on regional surface temperature trends in the context of global climate change.

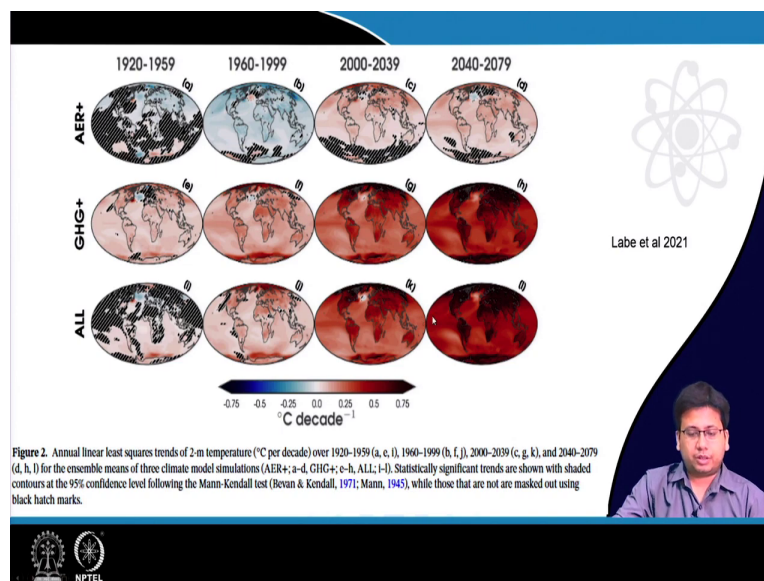
That is we know that the surface temperature is increasing that is not that is indisputable and it is also known that aerosols and greenhouse gases are causing this. But how exactly or what is the relative contributions of these etcetera that is not always clear and it is difficult to quantify.

Now, to address this issue, we use a new collection of initial condition large ensembles from the community earth system model version 1; this is the GCM that are prescribed with different combinations of industrial aerosol and greenhouse gas forcing. To compare the climate response to these external forcings, we adopt an artificial neural network architecture that from previous work that predicts the year by training on maps of near surface temperature.

That is like you give the near surface temperature map to the neural network and it will predict for you which year that belongs to.

So, and now, we then use layer-wise relevance propagation or to visualize the regional temperature signals that are important for the ANN's predictions in each climate model experiment.

(Refer Slide Time: 24:21)



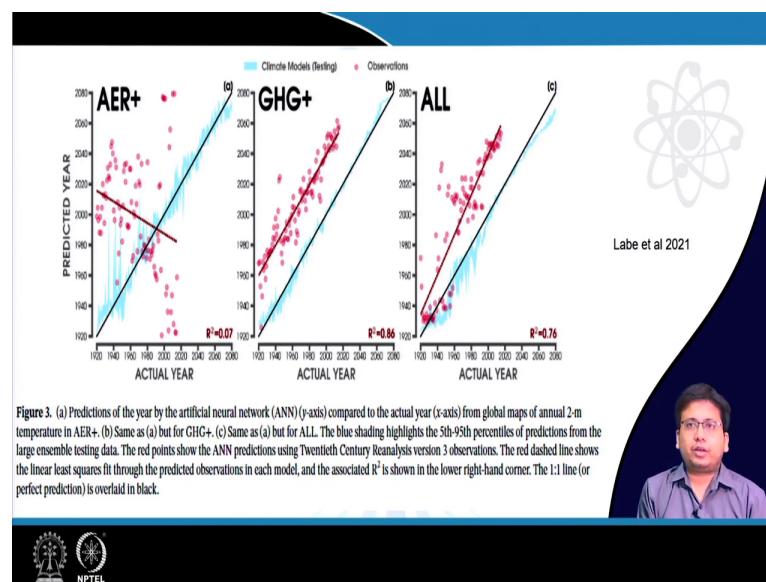
So, basically, the idea is as follows. So, like there is an earth system model like which they are mentioned as the CSM. So, that can be run under different settings. Let us say that we run it in one setting like where the aerosol is like there is a forcing of aerosol, keeping other forcing as constant. Let us say that there is no greenhouse gas emission; but the aerosol quantity keeps on increasing over the years. So, then, in such a setting, how will the global surface temperature looks like.

So, like in the different periods of time, this is the way it simulates and so, like we see that in this case, like if you consider the current period, the like we see that the heating has been relatively mild. Now, we consider another scenario, the greenhouse gases are emitted ah, but aerosols are not emitted. So, in that case, we see how the emission; I mean how that heating has taken place

over the years and in this case, we see that in different periods of time the heating has been much more significant or much more prominent.

And if both forcing years are used simultaneously, now this does not necessarily mean that the result will be this heating plus this heating. They can also be reduction also due to the coupling of the different variables. As I mentioned, increase of one variable may somehow offset or impede the growth of another variable and so on. It is difficult these things, these feedbacks are often very complex to model; but it turns out that if both are used simultaneously, then this is what the temperature evolution will take place.

(Refer Slide Time: 25:59)



Now, we have a neural network. What it does is it like it takes as input, the temperature map of any of the world for from any given year and tries to predict what that year is going to be. So, that machine learning model has already been trained on observations. So, let us say it is doing well. So, now, if that model is that machine learning model is now deployed in on each of the these scenarios say suppose this map is given to the to the neural network and it is asked to predict the year, then how will it be able to predict the year correctly or not.

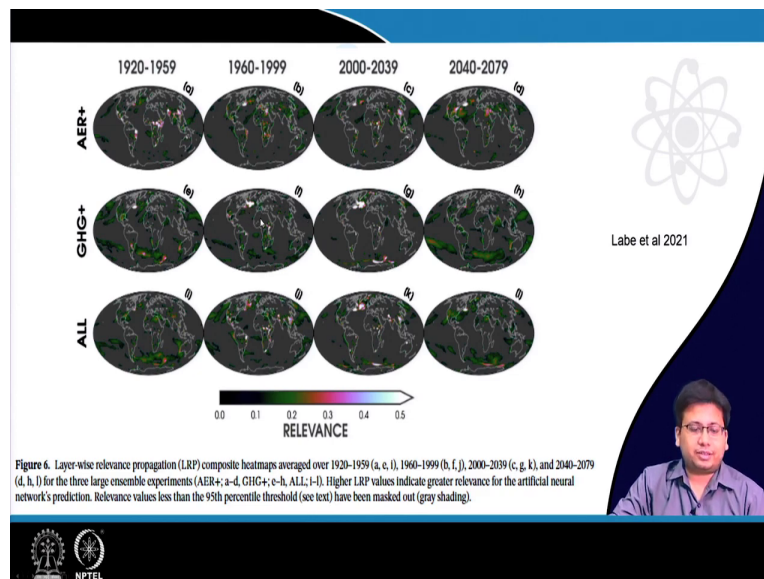
In this case also let us say we will do the same thing. We will give the temperature map of the world and ask it to predict the year and we will see whether it is able to predict or not. So, the

idea is like we will see that like if this is the in this scenario, like can we say that this temperature map is realistic or not; how close it is to realist reality or in this case, in the in this scenario, in this emission scenario, are we getting the realistic maps or not. Now, how do we understand whether a map is realistic?

The answer is if the trained neural network is able to predict the year correctly, then we understand that the map is realistic; else, it is not and it turns out that in case of greenhouse gases, it is actually able to predict the year reasonably, accurately with a small bias. But in the aerosol situation, then that is in the situation where there is only aerosol emission, in that case it is not able to predict the year that much accurately. I mean there is bigger bias in this case.

So, like as you can see the like in this case the  $R^2$  value is very high; that means, that the error is nearly equal to 0. But in this case the  $R^2$  is very low; that means, the error is quite high. But if both are considered, both kinds of emissions are considered, in that case we see that the  $R^2$  is again reasonably high.

(Refer Slide Time: 28:19)



Now, in each of the situations like that is given a temperature map, the this model is predicting the year. So, the question might arise that which regions is it focusing on to identify the year like from these emission maps. So, for that purpose, we use the layer wise relevance propagation. So,



you will as you may remember layer what layer wise relevance propagation does is that it tries to quantify the importance of every part of the input in predicting the output.

So, like it is like saying that if we have this temperature map and it is being predicted as say 1980 or something like that, then the neural network will focus on which regions to of the world to understand that like this is this year is indeed 1980. So, it turns out that these are the relevance of the difference different regions of the world.

So, I am not going into the details of what the relevance maps are; basically, the idea to be taken away is that like in each of the different scenarios, so in or in each emission scenarios, we see that only certain parts of the world are critical in understanding this matter. So, in other words, these are the crucial regions which are the drivers of climate change in this particular scenario.

So, the like using these approach of layer wise relevance propagation, if you want to understand which are the which factors or which variables, in which locations are the key drivers of climate change. In that case, this approach of layer wise relevance propagation might be a good way to go about and so, these are the different references we discussed today.

(Refer Slide Time: 30:12)



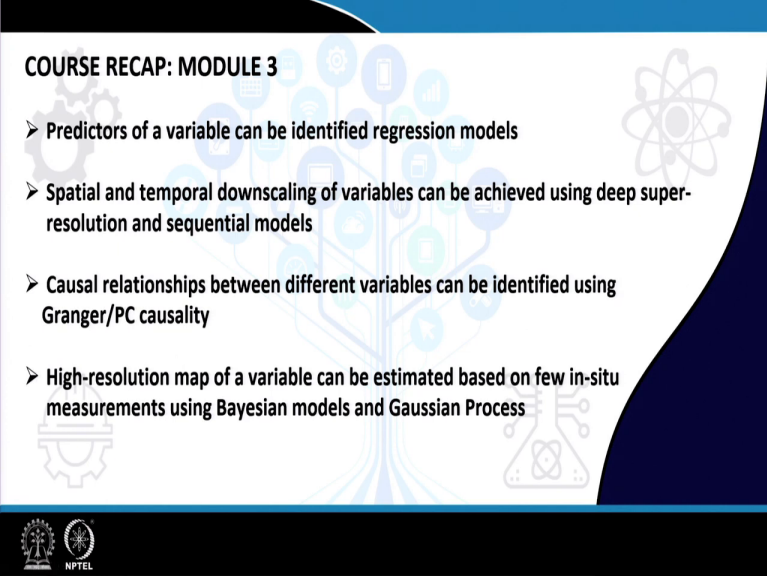
## REFERENCES

- Mansfield LA, Nowack PJ, Ksoar M, Everitt RG, Collins WJ, Voulgarakis A. Predicting global patterns of long-term climate change from short-term simulations using machine learning. *npj Climate and Atmospheric Science*. 2020 Nov 19;3(1):1-9.
- Labe ZM, Barnes EA. Detecting Climate Signals Using Explainable AI With Single-Forcing Large Ensembles. *Journal of Advances in Modeling Earth Systems*. 2021 Jun;13(6):e2021MS002464.
- Sun Z, Archibald AT. Multi-stage ensemble-learning-based model fusion for surface ozone simulations: A focus on CMIP6 models. *Environmental Science and Ecotechnology*. 2021 Oct 1;8:100124.





(Refer Slide Time: 30:19)



**COURSE RECAP: MODULE 3**

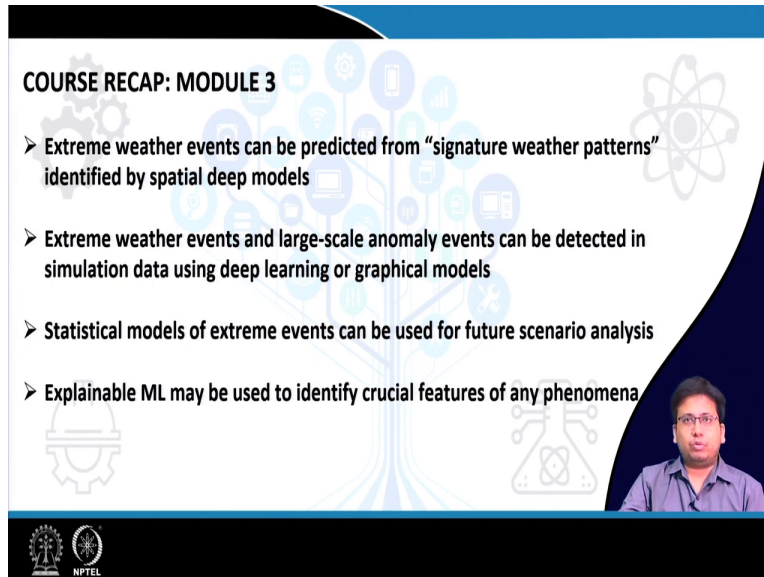
- Predictors of a variable can be identified regression models
- Spatial and temporal downscaling of variables can be achieved using deep super-resolution and sequential models
- Causal relationships between different variables can be identified using Granger/PC causality
- High-resolution map of a variable can be estimated based on few in-situ measurements using Bayesian models and Gaussian Process

The slide features a blue header and footer. The background is white with faint, stylized icons of a gear, a tree, a person, and a molecular structure. The bottom left corner contains the NPTEL logo.

So, now, we like since we have reached the end of this course, we do a brief recap of what we learnt in the different modules. So, in module 3, where we discussed like various knowledge discovery using the discovery of new knowledge about the earth system science using this like using machine learning. So, the takeaway messages are as follows. So, we saw that like for any particular variable, its predictors can be identified using regression model.

So, an example of that we saw in the predictors of Indian monsoon. Then, secondly, it is possible to do a spatial and temporal down scaling of variables using either deep super resolution models or using sequential models. It is also possible to find out the causal relationships between different variables using Granger causality with some difficulties and also, with PC based PC algorithm based causality. It is also possible to generate a high resolution map of a variable from a few in-situ measurements using Bayesian process models or Gaussian processes.

(Refer Slide Time: 31:27)



**COURSE RECAP: MODULE 3**

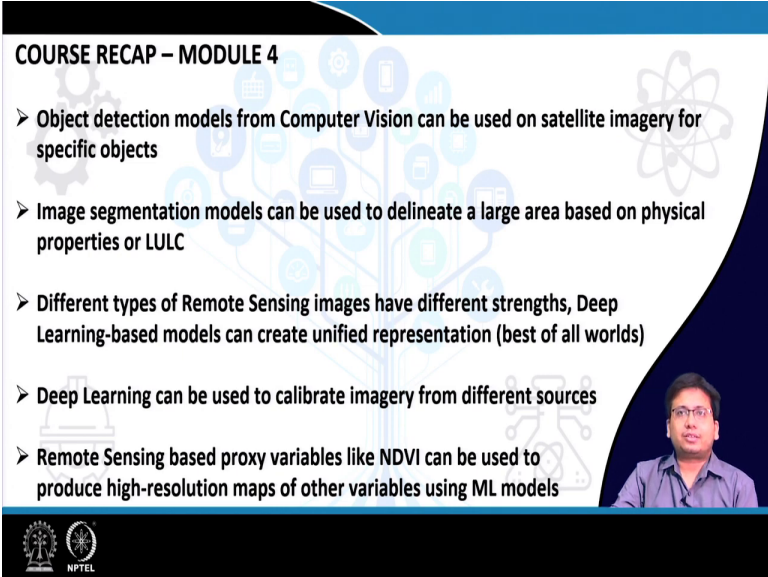
- Extreme weather events can be predicted from “signature weather patterns” identified by spatial deep models
- Extreme weather events and large-scale anomaly events can be detected in simulation data using deep learning or graphical models
- Statistical models of extreme events can be used for future scenario analysis
- Explainable ML may be used to identify crucial features of any phenomena

The slide features a background with various icons related to technology and science, including a tree-like structure of icons, a gear, a lightbulb, a smartphone, a laptop, a network diagram, and a molecular structure. A video inset in the bottom right corner shows a man with glasses speaking. The NPTEL logo is visible in the bottom left corner.

We also discussed about extreme weather events like we saw how extreme weather events like heat waves can be predicted a few days in advance from some signature weather patterns and these signature weather patterns can be identified by spatial deep models. We also saw how these extreme weather events or large scale anomaly events, they can be detected in vast volumes of simulation data using say either deep learning or graphical models.

We also learn to do like analysis of extreme events in future using the statistical models for extreme. And we also saw how explainable machine learning like this layer wise relevance propagation, backward optimization etcetera can be identified, used to identify the crucial features of any phenomena, just like what we discussed a few minutes back using to identifying the main drivers of climate change using layer wise relevance propagation.

(Refer Slide Time: 32:21)



**COURSE RECAP – MODULE 4**

- Object detection models from Computer Vision can be used on satellite imagery for specific objects
- Image segmentation models can be used to delineate a large area based on physical properties or LULC
- Different types of Remote Sensing images have different strengths, Deep Learning-based models can create unified representation (best of all worlds)
- Deep Learning can be used to calibrate imagery from different sources
- Remote Sensing based proxy variables like NDVI can be used to produce high-resolution maps of other variables using ML models

The slide features a background with faint icons of a gear, a smartphone, a laptop, and a network diagram. A small video inset in the bottom right corner shows a man with glasses and a blue shirt. The NPTEL logo is visible in the bottom left corner.

Then, we went to module 4, where we are discussing the applications of machine learning on earth observation systems. The key take away messages were that firstly, the like it is possible to use object detection models like these the RCNN and YOLO to these models can be used on satellite imagery to detect specific objects. Similarly, image segmentation models for computer vision can also be used to delineate like large area based on certain physical properties like water or like ground material etcetera or using land use land cover.

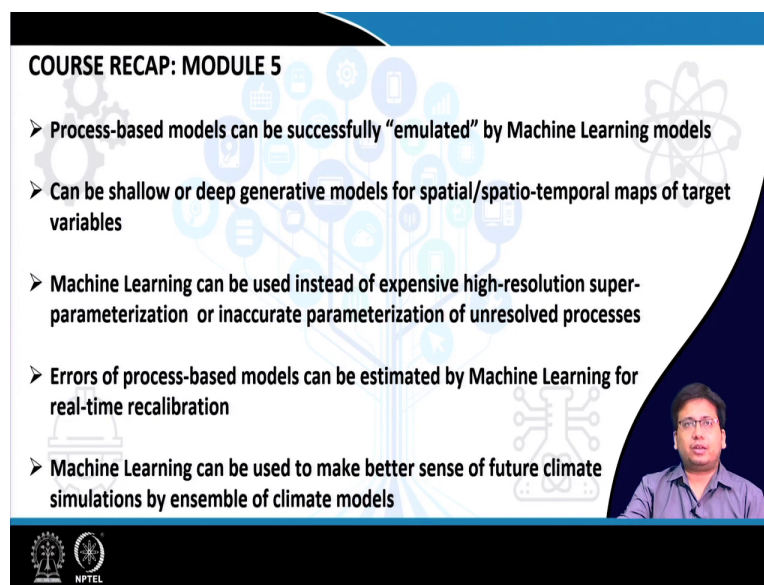
We also saw that different kinds of remote sensing images can have different strengths like some of them are may be may be having high spatial resolution, but low spectral resolution; some may have high spectral resolution, but low spatial resolution; some of them like maybe accurate very accurate, but spatially limited and so on and so forth.

Now, deep learning based models can be used to create a unified representation of all these kinds of model that can be like that can take the best features out of all and create a high resolution, high there is high spatial resolution, high spectral resolution and highly accurate models. Like the it can also be used to calibrate the imagery obtained from different sources.

That is like if different sources are measuring the same phenomena from different angles or using different technology, there will be discrepancies between them. But such discrepancies can be solved by with the help of deep learning by calibrating the different sources against each other.

And we also saw how the different remote sensing based proxy variables or indices like NDVI, how they can be used to produce high resolution spatial maps of other variables with the help of machine learning.

(Refer Slide Time: 34:29)



**COURSE RECAP: MODULE 5**

- Process-based models can be successfully “emulated” by Machine Learning models
- Can be shallow or deep generative models for spatial/spatio-temporal maps of target variables
- Machine Learning can be used instead of expensive high-resolution super-parameterization or inaccurate parameterization of unresolved processes
- Errors of process-based models can be estimated by Machine Learning for real-time recalibration
- Machine Learning can be used to make better sense of future climate simulations by ensemble of climate models

The slide features a background with faint icons of a gear, a lightbulb, a smartphone, and a network diagram. A small video inset in the bottom right corner shows a man with glasses speaking. The NPTEL logo is visible in the bottom left corner.

We saw a concrete example predicting ground water all over the world using NDVI. Next, we came to module 5, where we discussed how earth system processes process models can benefit from machine learning. So, we saw that the these process models can be successfully emulated by these machine learning models which means that like it is not really necessary to run these process-based models for which are very computationally expensive; instead, we can often just predict their outcomes using ML without actually running the models.

Like we also saw that like in some cases, it is possible to build very simple lightweight emulators of certain particular variables like without actually simulating them using a process-based model, like remember the stochastic weather generators and so on. So, those are like based on like very simple like those were not even based on deep learning, they are simply Bayesian models.

But like they can also be improved by taking into account deep generative models like variational auto encoders, gans etcetera. We saw some applications of Cgans, conditional gans and so on. We also saw how machine learning can be used to for re-parameterizations of machine learning models. That is it is no longer used to it is no longer necessary to couple the low resolution process models with high resolution and computationally expensive models at low resolution; instead machine learning can be used there.

Apart from that, like we also saw how the errors which are incurred by these process based models, how they can be corrected by a calibration at real time. So, we saw some examples of nudging and so on. So, that is a very useful tool and we also saw how machine learning can be used to make better sense of the future climate simulation by ensembles of climate models.

(Refer Slide Time: 36:38)



So, like this is what we discussed or the various aspects of this which we discussed in the earlier part of this lecture. So, now, if you are interested in further research and reading on these topics, so like you these are some of the different international conferences and journals, where you can find more papers. So, like there are there is this international conference of climate informatics which is held every year with the specific purpose of promoting research, where ML is used in climate science.

Apart from that the American Geophysical Union, European Geophysical Union, they hold their meetings every year. And like typically AGU meeting, it happens in December and EGU meetings happen in May and like they often have special sessions how machine learnings can be used in geosciences.

Apart from that, the top machine learning conferences they have their they often are hosting special workshops on how these ML methods can be used for the geosciences, for climate change mitigation etcetera and then, the different computer vision conferences, they also often have papers about how machine like how their methodology can be used say for example, in remote sensing and so on.

And apart from that, there are many journals from the Earth System Science domains like computers and geosciences is a good like is a good journal, apart from that journal of computer sciences. Then, as far as remote sensing is concerned, these that IEEE transactions on geosciences and remote sensing this is a very very useful journal. Then, Journal of Advances in Modelling of Earth Sciences.

So, as far as earth system process modelling is concerned, this is the best journal. Apart from that, other like domain specific journals like Geophysical Research Letters or Water Resources Research and some generic journals like nature communications, frontiers in science, frontiers in climate these are also very good venues.

So, with that, we come to the end of this lecture and this course. I hope you enjoyed this course and you learnt lot of new material and the I hope this will give you a broader perspective on how climate science or other aspects of geo science can benefit from machine learning model.

And whether you come from an ML background or whether you are coming from the geosciences background, I hope this course has given you a like a very new perspective and like given you a like if you are a climate or a if you are a geoscientist, I hope this has given you an idea of or the entire new toolbox for improving your domain.

And if you are coming from the machine learning background, I hope that you have been acquainted to an entire new set of applications, where your methodology can be useful. So, I

hope you will be using these your expertise to enrich this field. So, I will be looking forward to seeing great contributions from all of you in the future.

So, that brings us to the end of this course. Stay well and bye.