Dynamic Data Assimilation

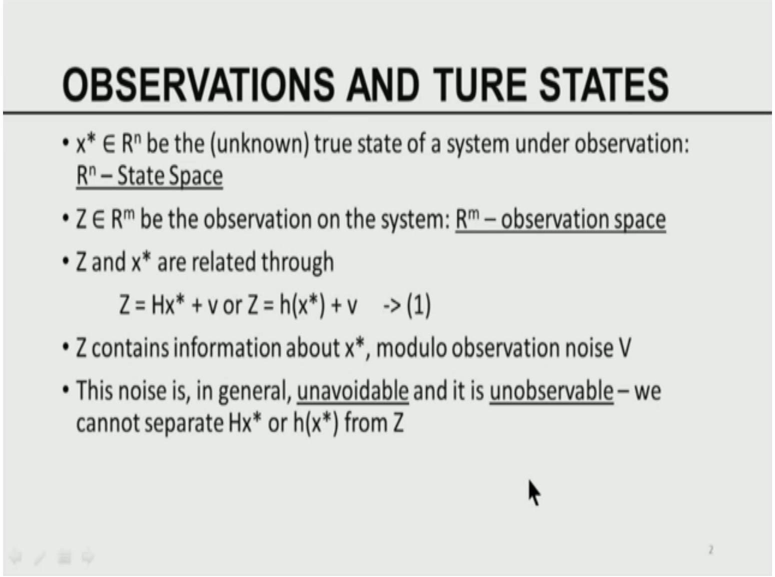
Prof. S Lakshmivarahan
School of Computer Science
Indian Institute of Technology, Madras

Lecture - 03
A classification of forecast error

In the last couple of lectures, we saw the role of data assimilation and it is importance to predictive science, prediction is forecast. We also saw yesterday, that not all the process is can be predicted accurately, some can be done rather precisely, in many cases forecast will not be perfect in imperfect forecast is said to have errors. So, we I am going to talk today's lecture with a good classification of forecast errors, because this classification will help us.

How do we attack the problem of correcting forecast errors using data assimilation and will also tell us what kind of errors need, what type of tools to be able to correct them in order to be able to make the forecast better. So, today's emphasis is going to be a classification of forecast errors and I would like to remind the reader that forecasting is fundamental aim of data assimilation and forecast errors are inherent in every forecast, in order to be able to correct the error. We need to have handle on the classification of errors.

(Refer Slide Time: 01:43)



OBSERVATIONS AND TURE STATES

- $x^* \in R^n$ be the (unknown) true state of a system under observation: R^n – State Space
- $Z \in R^m$ be the observation on the system: R^m – observation space
- Z and x^* are related through

  $Z = Hx^* + v$ or $Z = h(x^*) + v$ -> (1)
- Z contains information about x^*, modulo observation noise V
- This noise is, in general, unavoidable and it is unobservable – we cannot separate Hx^* or $h(x^*)$ from Z

To do that, I am going to start with the relation between the truth and the observation.

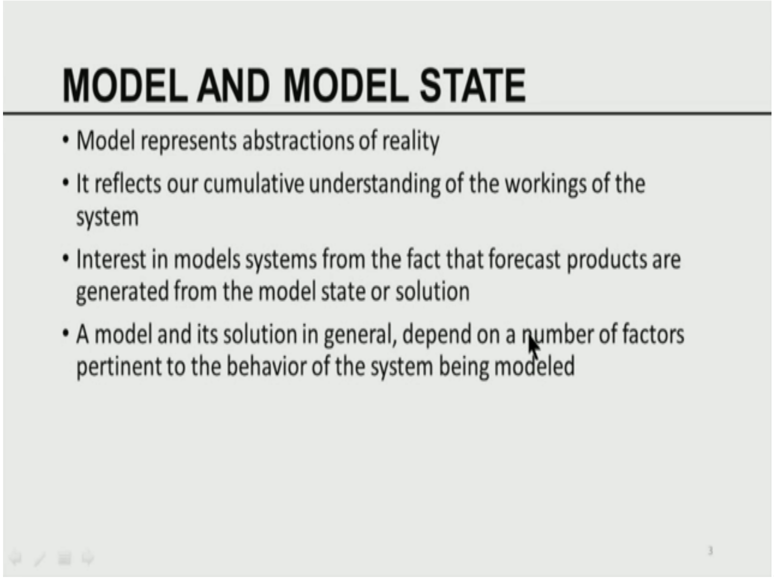
The truth is the true state of the Mother Nature, observations are data, obtained through sensing the Mother Nature’s true state. So, let us assume x star is a vector with the unknown true state of the system under observation for example, today's temperature in the city of London that is the true state of the Mother Nature, but we are going to observe the true state. So, Z is called the observation. Observation in general is a m vector, true state is n vector, the observation and the true state are related through a fundamental mathematical expression Z is equal to H of x star and v plus v. Here, v is the observation noise in the non-linear case, Z is equal to H of x star plus v H is a non-linear function.

So, observation may be related linearly to the true state or observation can be related to the two state non-linearly. In either case there are going to be errors correcting the observation. We are assuming the errors are additive in nature that is a simple way of dealing with observational errors and this aspect of considering observational errors being an addictive process has been around ever since the days of Gauss that we talked about in the last class. So, you can readily see if you want to know the true state of Mother Nature, you can only sense it through devices, the device output, the Z, the input to the devices are the x stars.

So, Z contains information about the true state of Mother Nature x star, but it is corrected by additive noise. So, we say Z contains the information modulo the observation noise v. This observation noise is in general unavoidable, it is also unobservable in what sense we may not, we will not be able to separate H of x star or H of x star from Z. If we are able to x separate H of x star from Z. We are able to have a filter, that will filter out the noise in general, such filtering is not easy to develop, because we may not know very precisely all the properties of the noise. We generally assume it is Gaussian distributed. It is also white and so on.

So, if want to know the true state of Mother Nature, you have to observe her evolution observation contains the secrets about Mother Nature and that is not unusual, when you feel not too well, you go to the doctor, the doctor wants to be able estimate your true state of the physical system, the true state of the physical system are obtained by making observation, blood pressure temperature, various kinds of tests and so, on. So, observations are indicators of the underlying true state of any system beat human or nature.

(Refer Slide Time: 05:34)



MODEL AND MODEL STATE

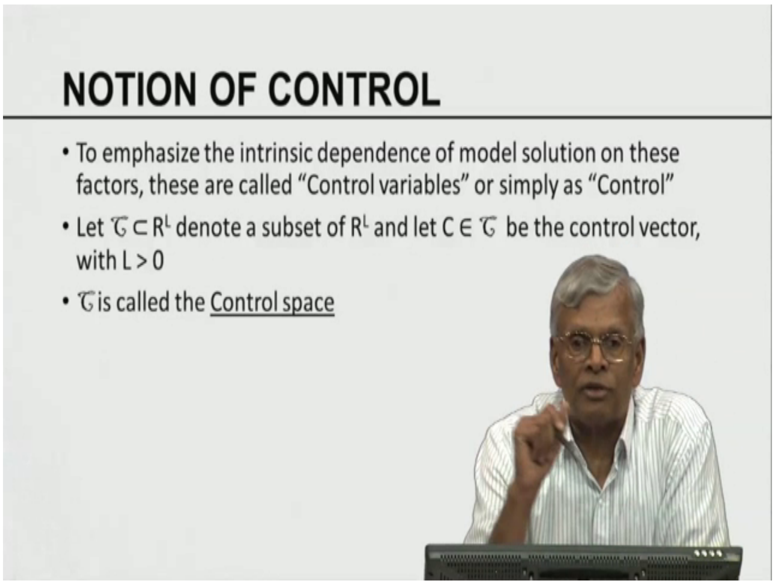
- Model represents abstractions of reality
- It reflects our cumulative understanding of the workings of the system
- Interest in models systems from the fact that forecast products are generated from the model state or solution
- A model and its solution in general, depend on a number of factors pertinent to the behavior of the system being modeled

.

Now let us pull the other one, we are talked about model, what are the model? Models represent abstraction of reality, models represent our understanding of how Mother Nature works. It reflects our cumulative understanding of the working of the system, interesting model stems from must be stems from not systems. I am sorry, stems from the fact that the forecast product are generated from model state our solutions. So, to do forecast, models are necessary, models represent our understanding of the Mother Nature, our understanding, the Mother Nature sometimes closer to being perfect sometimes may not be perfect, a model and it is solution in general depend on number of factors pertinent to the behaviour of the system being modelled. We have already talked about the role of parameters in models and so, on.

So, now, here comes the two facets, the reality as it is our sense of our sensing of reality in terms of observation, models represent our understanding of our reality, probably words and that is probably the gap between the two. It is this gap between the actual reality and our understanding of how reality works leads to forecast errors, if the model is perfect the forecasts are perfect. If the model and the reality, if there is a gap that gap reflects in the form of forecast errors. Now, I would like to be able to classify the presence of this gap between our understanding of reality and actual reality itself.

(Refer Slide Time: 07:21)



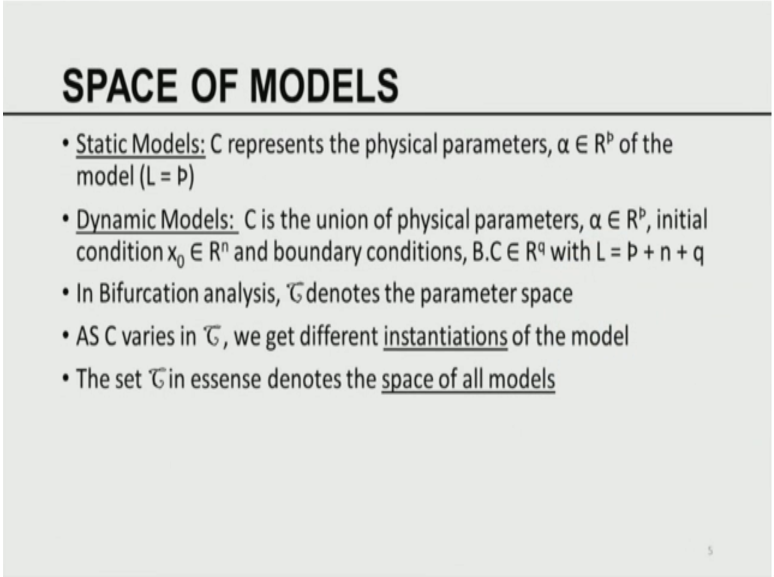
NOTION OF CONTROL

- To emphasize the intrinsic dependence of model solution on these factors, these are called “Control variables” or simply as “Control”
- Let $\mathcal{C} \subset R^L$ denote a subset of R^L and let $C \in \mathcal{C}$ be the control vector, with $L > 0$
- $\mathcal{C}$ is called the <u>Control space</u>

To emphasis the intrinsic dependence of model solution on various factors. So, we have already seen, if it is a dominate model the model, the solution depends on the initial condition parameters boundary conditions. So, the model solutions are contingent on the value that we are saying to these variables, because these variables control the model solution, anybody who has done anything in differential equation knows the differential equations solution. I have a general solution you get a particular solution by specializing the initial conditions, if we change the initial condition the solution changes. So, changing the initial condition changes a solution. In other words initial condition controls the evolution of the solution, change in the parameters controls the evolutions of the solution. So, anything that can change the model solution is called a control variable. So, control in principle refers to all the factors collectively that affects the evolution of the model solution based static parametric module.

Let's C refers to a subset of R L; R L; L is a integer R L is a space of real vector of size n. I am assuming C is a subset of R L; that means, any vector C belonging to script C is the control vector of dimension, L total the real number of control in general L C is called the control space every point in control space corresponds to one solution of the model, if you change the control vector the model solution changes. So, ultimately the behaviour of the forecast depends on the value of the control that we use and I would like to quickly remind that the control consist of initial condition, boundary condition parameters, anything that is part of the model, if I change any one of these factors of the solution changes, I call it control.

(Refer Slide Time: 09:32)



SPACE OF MODELS

- Static Models: C represents the physical parameters, $\alpha \in R^p$ of the model (L = Þ)
- Dynamic Models: C is the union of physical parameters, $\alpha \in R^p$, initial condition $x_0 \in R^n$ and boundary conditions, B.C $\in R^q$ with L = Þ + n + q
- In Bifurcation analysis, $\mathcal{C}$ denotes the parameter space
- AS C varies in $\mathcal{C}$, we get different instantiations of the model
- The set $\mathcal{C}$ in essense denotes the space of all models

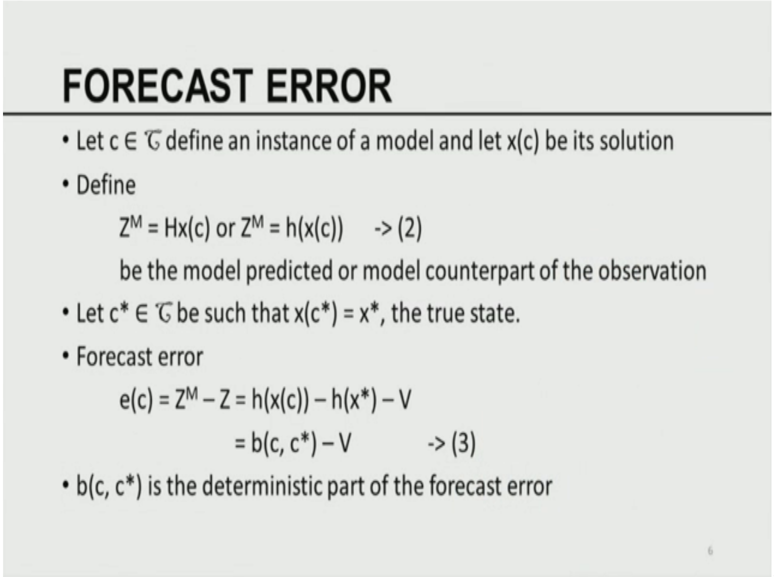
In static model control represents only the physical parameters in a static model, there is no initial condition, there is no boundary condition, it is a bunch of parameters and that is all. So, in that case the parameters we call it alpha. Alpha is a p vector, in this case the l is equal to p the control space is essentially a parameter space in general, I want you to remember the parameter space is only a subset of the control space, the control space consists of initial conditions, boundary conditions and parameters, but parameters are a parameter set, a set is a subset of all the controls. So, in the static model there is no initial conditions, there is no boundary conditions simply parameters.

The dynamic model, it is the control, is a union of parameters initial condition boundary conditions. So, l is equal to p plus n plus q where is p coming from p is the number of parameters, n is the number of initial condition, q is the number of boundary condition. So, l is equal to p plus the n plus q in non-linear differential equation. There is a branch called bifurcation analysis. In bifurcation analysis c represents the parameter space, the bifurcation analysis depends on variation of the behaviour with respect to variation of parameters in the parameter space. So, as c varies in script c. we get different instantiations of the model anybody who knows differential equation knows that if the initial condition changes it represents a different model the parameter changes is represent different model. So, each model within a class. So, for a particular choice of the parameters, we call it an instant of that model, the instant being picked by the valves are controlled.

The set c in a sense denotes the set of all models. So, from a model now we are considering, not one model, a class of models. So, in general in science a model does not mean one model, a model means a class of model when we say primitive equation model, primitive equation model is not one, but it is an infinite collection of models Barotropic Vorticity equation model is not one, but the collection model shallow water model, likewise same thing with respect to harmonic oscillator. Harmonic oscillators are a very generic thing, the frequency if you change it, the model changes, initial condition changes. It changes if you add a friction, it changes if you add a forcing it changes. So, by model we always mean infinite class.

I have to pick from this infinite class, a particular model that can be utilized the picking of the particular model, from the class is done by specifying the values of the control, the control consists of parameters, initial conditions, boundary condition, whatever applies whether it is dynamic or static.

(Refer Slide Time: 12:24)



FORECAST ERROR

- Let $c \in \mathcal{C}$ define an instance of a model and let $x(c)$ be its solution
- Define

  $Z^M = Hx(c)$ or $Z^M = h(x(c))$ -> (2)

  be the model predicted or model counterpart of the observation
- Let $c^* \in \mathcal{C}$ be such that $x(c^*) = x^*$, the true state.
- Forecast error

  $e(c) = Z^M - Z = h(x(c)) - h(x^*) - V$

  $= b(c, c^*) - V$ -> (3)
- $b(c, c^*)$ is the deterministic part of the forecast error

.

So, now with this as a background I am now going to define the classification of forecast errors. Let c be in instance of the model. So, in other words I am representing a model by the choice of control vector. So, if c is the control vector x c, let it denote the solution a c varies x c varies. So, define z superscript M super Z super script m. So, what does this mean x e is the model output, h is the operator Z of M is the model counterpart of the observation Z of M, the non-linear case model counterpart of the observation.

Now, I would like to distinguish between model counterpart or model predicted observation from the actual observation z is the actual observation, comes from the meter that I read satellite radar voltmeter ammeter whatever is, but if I know the state, if I know h, I can also predict what the model predicted observation at today. So, there are two versions of the observation, the actual observation, the model predicted observation. So, let's c star b, the c such that x of c star is x star, the true state if I assume the model is perfect, my model includes perfect model. A perfect model has to be parameterized, let c star be the parameter that gives me the perfect model I do not, what c star is, but I am assuming, such that c star exists.

If your modelling is good such as c star exists, there's no question about the existence of that. So, this is where the whole thing lies, there is a c star that corresponds to the true state of nature the model matches the Mother Nature, but there is a c. I have picked the c may not be c star, if the c is not equal to c star the x of c is not equal to x of c star x of c star corresponds to the true state of nature x of c is the state predicted by the model that corresponds to parameter c and these two in general need not to be the same and that's where fundamentally forecast errors arise.

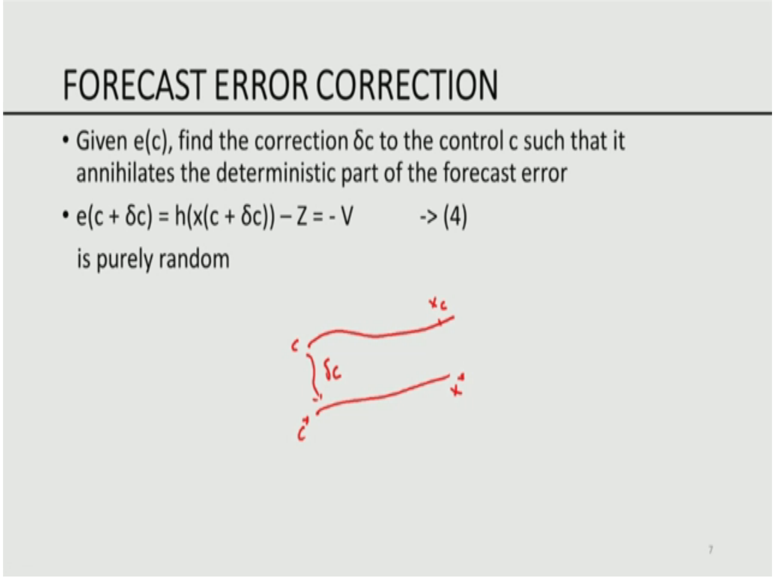
So, the forecast errors now, can be defined as error in the model induced by the control vector c e of c is equal to z super m what is that that is the model counterpart of the error this is generated by z the z without any superscript that is the true state of Mother Nature the difference between the two is what i just talked about what the model sees what the Mother Nature has the difference between the two is called the forecast error. So, the forecast errors now has a description z m is a h of x c z is equal h of x star plus v. So, there you get this following equation 4 the first term I am going to call it b of c comma c star minus v; v is the observation noise what is this b; b is the deterministic part of the forecast error.

So, the forecast errors consist of two parts one due to the unavoidable unobservable random error v, the second one the deterministic part, which is which arises largely because of my inadequate knowledge about what Mother Nature does, she uses c star I use c you see its not equal to c star there is going to be a bias. So, you can think of b as a bias in the in the forecast the bias is a function of c and c star.

So, this is the general framework, in other words the forecast errors always consists of two parts a deterministic part and a random part. I also want to quickly add random part. We cannot touch it we cannot annihilate, the random part stays with the observation. So, what is the best you can do if you want to be able to reduce the forecast error? The only thing that you can do is to hope to annihilate b, if you can annihilate b then you will be left to only with the random errors which is uncontrollable.

So, what is the basic idea of forecast error classification I would like to be able to understand what part of the forecast error I can control? What part of the forecast error I cannot control? We can only deal with things that I can control over do not worry about things that you have no control over. So, the separation of the forecast error into deterministic part and the stochastic part is very helpful in trying to design schemes, for correcting forecast errors that is the motivation for this.

(Refer Slide Time: 17:44)



FORECAST ERROR CORRECTION

- Given e(c), find the correction δc to the control c such that it annihilates the deterministic part of the forecast error
- $e(c + \delta c) = h(x(c + \delta c)) - Z = - V$ -> (4)

  is purely random

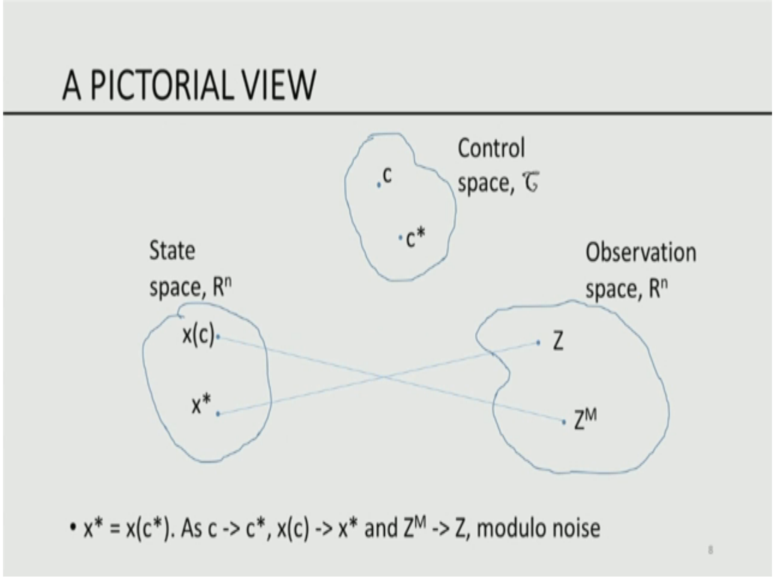
So, given e of c. Now, please go back e of c is given by equation 3. So, you can think of e is equal to b minus v. So, given e of c. So, what is that we would like to be able to do what is the concept of forecast error control there is c star. So, we have talked about separation of forecast errors in the deterministic part and the random part. Now, we are going to look at a classification of forecast errors, we know e c is the forecast error e c.

So, what is the basic idea here is c using c, I am going to generate a solution, I am going to get a forecast. Let us say x k or I will simply say x of c x of c, but I have x star which is the unknown to state x star is different from x c.

So, the question is how do I change x c to x star we all know x star depends on c star. So, the only way to move x c to x star is to change c to c star and that can be done by adding a perturbation delta c a that is what is being talked about in here. So, if you want to be able to annihilate the error you have to be able to change the control, you have to add a correction delta c to c and if c plus delta c is equal to c star, it will become h of c x of c star and that will be the true system or the true state and z represents the truth. So, the truth minus truth cancels itself and v is the uncontrollable unavoidable noise in this case this is purely random. So, if you look at fundamentally, how do we can improve the quality of forecast herein lies the solution.

The only way to be able to improve the quality of forecast is to find an increment delta c to the control, which in added to the control c will annihilate the deterministic part of the forecast error that is the fundamental relation that one needs to bear.

(Refer Slide Time: 20:39)



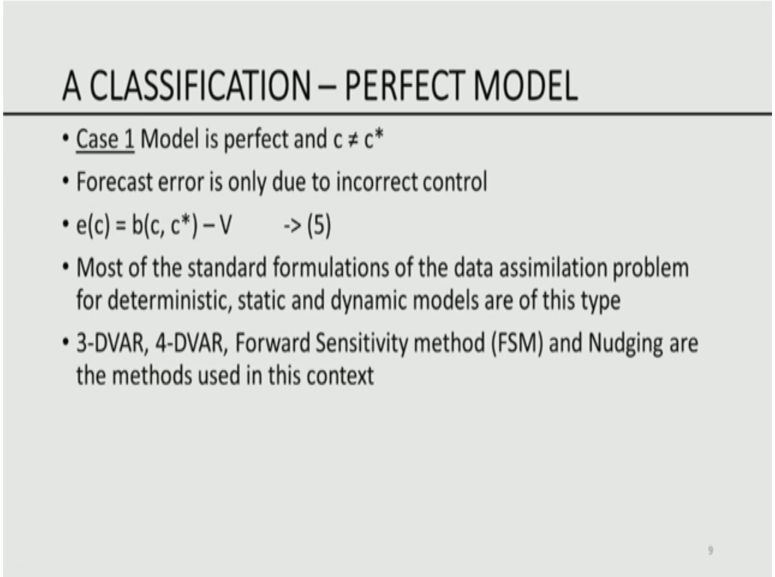
A PICTORIAL VIEW

- $x^* = x(c^*)$. As $c \to c^*$, $x(c) \to x^*$ and $Z^M \to Z$, modulo noise

So, this can be pictorially represented like this. So, the control space c represent the current belief about the model, c star is the unknown truth, if I use the c, I have picked a model x of c, if I picked x f c; x f c give me the observation z of m, but c star has x star that gives observation z. So, how do I? How do I minimize the difference between z and z star? In order to be able to minimize the difference between z and z star I should be able to minimize the difference between c and c star that is where the control lies. So, I what is the increment, I should add to the control in order to force x c closer to x star which will in turn make Z closer to Z M. So, this pictorial view is the basis for classification forecast errors.

So, I see. So, look at this mathematically, now I see tend towards c star x of c will tend towards x star, which will then imply Z of M will tend toward Z when Z of M tend toward z means what my model reflects Mother Nature, I cannot do any difference, any better that. So, this picture essentially tells you how one can hope to control the forecast error largely due to the difference between the model forecast and the true state of the system.

(Refer Slide Time: 22:19)



A CLASSIFICATION – PERFECT MODEL

- Case 1 Model is perfect and $c \neq c^*$
- Forecast error is only due to incorrect control
- $e(c) = b(c, c^*) - V$ -> (5)
- Most of the standard formulations of the data assimilation problem for deterministic, static and dynamic models are of this type
- 3-DVAR, 4-DVAR, Forward Sensitivity method (FSM) and Nudging are the methods used in this context

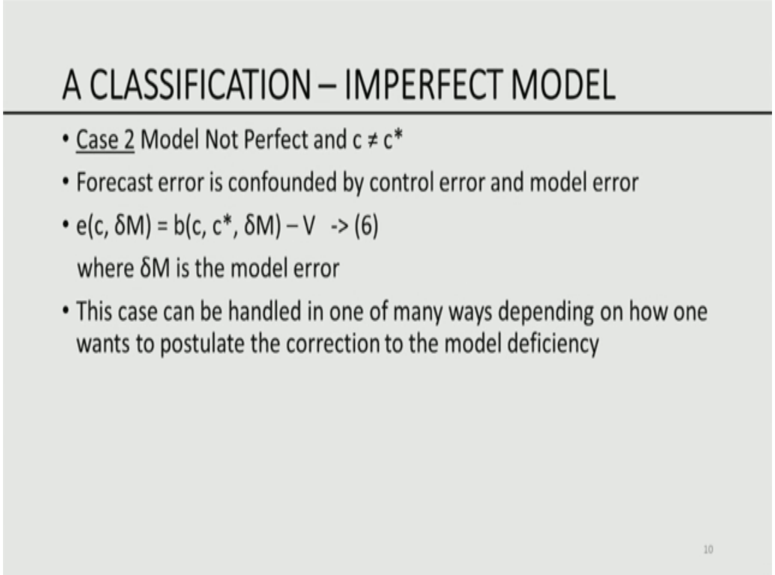
So, with that as the basis and now I am going to provide the actual classification. We have been talking about case one; in this case model is perfect, model is perfect means I have thorough understanding of Mother Nature, but I did not pick my c to b equal to c star, I may have a total understanding in the process, but I may not know the initial state of the Mother Nature. So, c is not equal to c star.

So, in this case the forecast error largely due to incorrect control, the model is capable of replicating Mother Nature, but I did not know the actual parameter, Mother Nature uses I only. Guess it, c is my guess, c star is her choice. The difference between c and c star is going to reflect this difference between c and c star reflects the forecast error. So, e of c in this case be c; c star minus v almost all the standard formulations of data simulation problem, for deterministic static and dynamic models are of this type. So, what does it be assume? We assume my model is perfect.

Let us talk about that for a moment. Now, no modeller believes that model is not, correct because if it is not correct we will not use it. So, if I am going to use the biotropic watch as an equation to be able to describe the hurricane scientists know that it captures 90 95 percent of reality is very close to being perfect. If it does not, if scientist does not have that confidence in the model they will not use it. So, much of the development in the forecast literature fall into this category namely models are perfect. We assume the models to be perfect, even the perfect model if that is going be the forecast error is largely due to the difference in control, if there is a forecast error only, because c is not equal to c star. I have the ability to be able to change the control thereby force the forecast error to be purely random.

So, most of the standard formulations of data assimilation falls in the category, the well-known 3 D-VAR, 4 D-VAR, forward sensitive method, nudging are all some of the examples of methods used to do data assimilation fall under this category.

(Refer Slide Time: 24:50)



A CLASSIFICATION – IMPERFECT MODEL

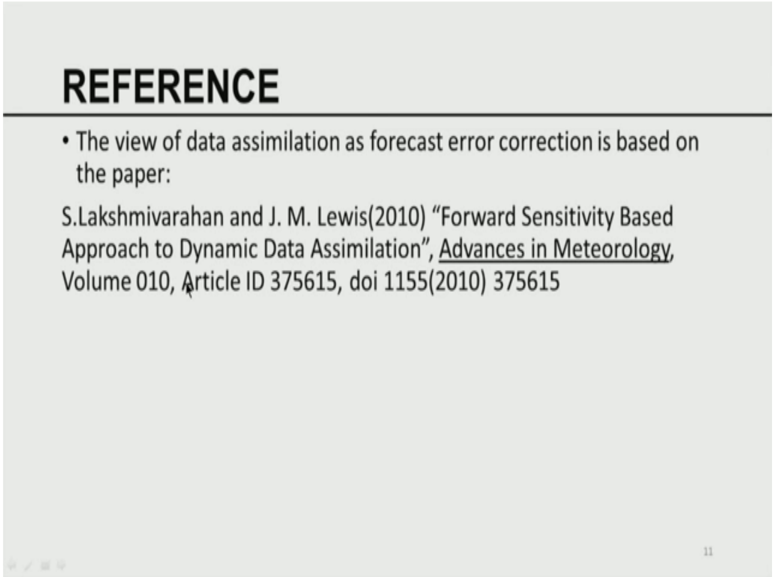
- Case 2 Model Not Perfect and $c \neq c^*$
- Forecast error is confounded by control error and model error
- $e(c, \delta M) = b(c, c^*, \delta M) - V \;\; \text{->} \; (6)$
  where δM is the model error
- This case can be handled in one of many ways depending on how one wants to postulate the correction to the model deficiency

10

Case two; this is much more difficult case the model is imperfect if them. So, is a model is imperfect and my control is not the same. So, there is two kinds of errors one coming from the model not being perfect, another coming from the fire control is not perfect. So, there are two types of errors that are confounded, it is very these confounded errors are very difficult to separate, we cannot say this part of the error is due to this part of the error, due to this confounding is a large headache.

The forecast error is the confounding of the model error and the control error. In this case we need much more powerful techniques and this is the most difficult case that one can deal with this case can be handled in one of many ways depending on how one wants to postulate the correction to the model error efficiency. In other words you have to want to correct the model error in a particular way, the way that you would like to be able to correct the model error will dictate the method by which you are going to correct the forecast error.

(Refer Slide Time: 26:12)



REFERENCE

- The view of data assimilation as forecast error correction is based on the paper:

S.Lakshmivarahan and J. M. Lewis(2010) "Forward Sensitivity Based Approach to Dynamic Data Assimilation", Advances in Meteorology, Volume 010, Article ID 375615, doi 1155(2010) 375615

So, this view of data assimilation as a forecast error, correction was proposed in a paper by Lakshmivarahan and Lewis in 2010. It is a basis for the paper forwards sensitivity based approach to dynamic data assimilation. These three error classifications have been the subject of this paper and the forwards sensitivity method, they had proposed is one of the methods by which we can correct model errors as in case one, with that we have concluded analysis of classification of forecast errors.

Thank you.